Yangon University of Economics

Post Graduate Diploma in Research Studies

(9th Batch)

DRS-113 Probability Theory

First Quarter

4. Probability & Counting Rules

5. Discrete Probability distribution

6. The Normal distribution

4

Probability and Counting Rules

STATISTICS TODAY

Would You Bet Your Life?

Humans not only bet money when they gamble, but also bet their lives by engaging in unhealthy activities such as smoking, drinking, using drugs, and exceeding the speed limit when driving. Many people don't care about the risks involved in these activities since they do not understand the concepts of probability. On the other hand, people may fear activities that involve little risk to health or life because these activities have been sensationalized by the press and media.

In his book *Probabilities in Everyday Life* (Ivy Books, p. 191), John D. McGervey states

> *When people have been asked to estimate the frequency of death from various causes, the most overestimated categories are those involving pregnancy, tornadoes, floods, fire, and homicide. The most underestimated categories include deaths from diseases such as diabetes, strokes, tuberculosis, asthma, and stomach cancer (although cancer in general is overestimated).*

The question then is, Would you feel safer if you flew across the United States on a commercial airline or if you drove? How much greater is the risk of one way to travel over the other? See Statistics Today—Revisited at the end of the chapter for the answer.

In this chapter, you will learn about probability—its meaning, how it is computed, and how to evaluate it in terms of the likelihood of an event actually happening.

OUTLINE

OBJECTIVES

After completing this chapter, you should be able to:

1. Determine sample spaces and find the probability of an event, using classical probability or empirical probability.
2. Find the probability of compound events, using the addition rules.
3. Find the probability of compound events, using the multiplication rules.
4. Find the conditional probability of an event.
5. Find the total number of outcomes in a sequence of events, using the fundamental counting rule.
6. Find the number of ways that *r* objects can be selected from *n* objects, using the permutation rule.
7. Find the number of ways that *r* objects can be selected from *n* objects without regard to order, using the combination rule.
8. Find the probability of an event, using the counting rules.

Introduction

A cynical person once said, "The only two sure things are death and taxes." This philosophy no doubt arose because so much in people's lives is affected by chance. From the time you awake until you go to bed, you make decisions regarding the possible events that are governed at least in part by chance. For example, should you carry an umbrella to work today? Will your car battery last until spring? Should you accept that new job?

Probability as a general concept can be defined as the chance of an event occurring. Many people are familiar with probability from observing or playing games of chance, such as card games, slot machines, or lotteries. In addition to being used in games of chance, probability theory is used in the fields of insurance, investments, and weather forecasting and in various other areas. Finally, as stated in Chapter 1, probability is the basis of inferential statistics. For example, predictions are based on probability, and hypotheses are tested by using probability.

The basic concepts of probability are explained in this chapter. These concepts include *probability experiments, sample spaces,* the *addition* and *multiplication rules,* and the *probabilities of complementary events.* Also in this chapter, you will learn the rule for counting, the differences between permutations and combinations, and how to figure out how many different combinations for specific situations exist. Finally, Section 4–5 explains how the counting rules and the probability rules can be used together to solve a wide variety of problems.

4–1 Sample Spaces and Probability

The theory of probability grew out of the study of various games of chance using coins, dice, and cards. Since these devices lend themselves well to the application of concepts of probability, they will be used in this chapter as examples. This section begins by explaining some basic concepts of probability. Then the types of probability and probability rules are discussed.

Basic Concepts

Processes such as flipping a coin, rolling a die, or drawing a card from a deck are called *probability experiments.*

OBJECTIVE 1

Determine sample spaces and find the probability of an event, using classical probability or empirical probability.

A **probability experiment** is a chance process that leads to well-defined results called outcomes.

An **outcome** is the result of a single trial of a probability experiment.

A trial means flipping a coin once, rolling one die once, or the like. When a coin is tossed, there are two possible outcomes: head or tail. (*Note:* We exclude the possibility of a coin landing on its edge.) In the roll of a single die, there are six possible outcomes: 1, 2, 3, 4, 5, or 6. In any experiment, the set of all possible outcomes is called the *sample space.*

A **sample space** is the set of all possible outcomes of a probability experiment.

Some sample spaces for various probability experiments are shown here.

Experiment	Sample space
Toss one coin	Head, tail
Roll a die	1, 2, 3, 4, 5, 6
Answer a true/false question	True, false
Toss two coins	Head-head, tail-tail, head-tail, tail-head

It is important to realize that when two coins are tossed, there are *four* possible outcomes, as shown in the fourth experiment above. Both coins could fall heads up. Both coins could fall tails up. Coin 1 could fall heads up and coin 2 tails up. Or coin 1 could fall tails up and coin 2 heads up. Heads and tails will be abbreviated as H and T throughout this chapter.

EXAMPLE 4–1 Rolling Dice

Find the sample space for rolling two dice.

SOLUTION

Since each die can land in six different ways, and two dice are rolled, the sample space can be presented by a rectangular array, as shown in Figure 4–1. The sample space is the list of pairs of numbers in the chart.

FIGURE 4–1
Sample Space for Rolling Two Dice (Example 4–1)

	Die 2					
Die 1	1	2	3	4	5	6
1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

EXAMPLE 4–2 Drawing Cards

Find the sample space for drawing one card from an ordinary deck of cards.

SOLUTION

Since there are 4 suits (hearts, clubs, diamonds, and spades) and 13 cards for each suit (ace through king), there are 52 outcomes in the sample space. See Figure 4–2.

FIGURE 4–2 Sample Space for Drawing a Card (Example 4–2)

A	2	3	4	5	6	7	8	9	10	J	Q	K
♥	♥	♥	♥	♥	♥	♥	♥	♥	♥	♥	♥	♥
A	2	3	4	5	6	7	8	9	10	J	Q	K
♦	♦	♦	♦	♦	♦	♦	♦	♦	♦	♦	♦	♦
A	2	3	4	5	6	7	8	9	10	J	Q	K
♠	♠	♠	♠	♠	♠	♠	♠	♠	♠	♠	♠	♠
A	2	3	4	5	6	7	8	9	10	J	Q	K
♣	♣	♣	♣	♣	♣	♣	♣	♣	♣	♣	♣	♣

EXAMPLE 4–3 Gender of Children

Find the sample space for the gender of the children if a family has three children. Use B for boy and G for girl.

SOLUTION

There are two genders, boy and girl, and each child could be either gender. Hence, there are eight possibilities, as shown here.

BBB BBG BGB GBB GGG GGB GBG BGG

In Examples 4–1 through 4–3, the sample spaces were found by observation and reasoning; however, another way to find all possible outcomes of a probability experiment is to use a *tree diagram.*

A **tree diagram** is a device consisting of line segments emanating from a starting point and also from the outcome point. It is used to determine all possible outcomes of a probability experiment.

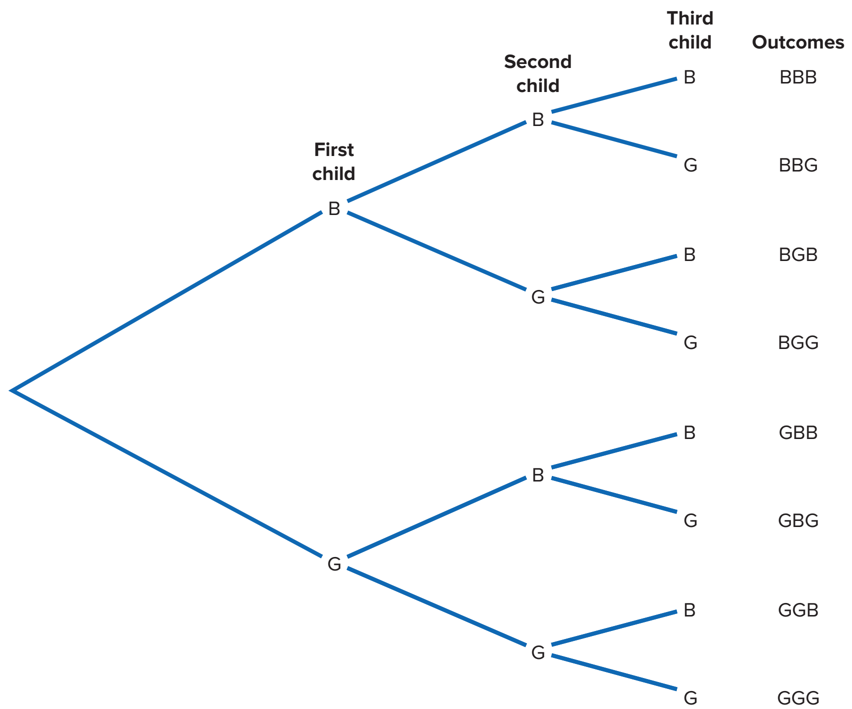
EXAMPLE 4–4 Gender of Children

Use a tree diagram to find the sample space for the gender of three children in a family, as in Example 4–3.

SOLUTION

Since there are two possibilities (boy or girl) for the first child, draw two branches from a starting point and label one B and the other G. Then if the first child is a boy, there are two possibilities for the second child (boy or girl), so draw two branches from B and label one B and the other G. Do the same if the first child is a girl. Follow the same procedure for the third child. The completed tree diagram is shown in Figure 4–3. To find the outcomes for the sample space, trace through all the possible branches, beginning at the starting point for each one.

FIGURE 4–3 Tree Diagram for Example 4–4



Historical Note

The famous Italian astronomer Galileo (1564–1642) found that sums of 10 and 11 occur more often than any other sum when three dice are tossed. Previously, it was thought that a sum of 9 occurred more often than any other sum.

Historical Note

A mathematician named Jerome Cardan (1501–1576) used his talents in mathematics and probability theory to make his living as a gambler. He is thought to be the first person to formulate the definition of classical probability.

An outcome was defined previously as the result of a single trial of a probability experiment. In many problems, one must find the probability of two or more outcomes. For this reason, it is necessary to distinguish between an outcome and an event.

An **event** consists of a set of outcomes of a probability experiment.

Historical Note

During the mid-1600s, a professional gambler named Chevalier de Méré made a considerable amount of money on a gambling game. He would bet unsuspecting patrons that in four rolls of a die, he could get at least one 6. He was so successful at the game that some people refused to play. He decided that a new game was necessary to continue his winnings. By reasoning, he figured he could roll at least one double 6 in 24 rolls of two dice, but his reasoning was incorrect and he lost systematically. Unable to figure out why, he contacted a mathematician named Blaise Pascal (1623–1662) to find out why.

Pascal became interested and began studying probability theory. He corresponded with a French government official, Pierre de Fermat (1601–1665), whose hobby was mathematics. Together the two formulated the beginnings of probability theory.

An event can be one outcome or more than one outcome. For example, if a die is rolled and a 6 shows, this result is called an *outcome,* since it is a result of a single trial. An event with one outcome is called a **simple event.** The event of getting an odd number when a die is rolled is called a **compound event,** since it consists of three outcomes or three simple events. In general, a compound event consists of two or more outcomes or simple events.

There are three basic interpretations of probability:

1. Classical probability
2. Empirical or relative frequency probability
3. Subjective probability

Classical Probability

Classical probability uses sample spaces to determine the numerical probability that an event will happen. You do not actually have to perform the experiment to determine that probability. Classical probability is so named because it was the first type of probability studied formally by mathematicians in the 17th and 18th centuries.

Classical probability assumes that all outcomes in the sample space are equally likely to occur. For example, when a single die is rolled, each outcome has the same probability of occurring. Since there are six outcomes, each outcome has a probability of $\frac{1}{6}$. When a card is selected from an ordinary deck of 52 cards, you assume that the deck has been shuffled, and each card has the same probability of being selected. In this case, it is $\frac{1}{52}$.

Equally likely events are events that have the same probability of occurring.

Formula for Classical Probability

The probability of any event E is

$$\frac{\text{Number of outcomes in } E}{\text{Total number of outcomes in the sample space}}$$

This probability is denoted by

$$P(E) = \frac{n(E)}{n(S)}$$

where $n(E)$ is the number of outcomes in E and $n(S)$ is the number of outcomes in the sample space S.

Probabilities can be expressed as fractions, decimals, or—where appropriate—percentages. If you ask, “What is the probability of getting a head when a coin is tossed?” typical responses can be any of the following three.

“One-half.”

“Point five.”

“Fifty percent.”[1]

These answers are all equivalent. In most cases, the answers to examples and exercises given in this chapter are expressed as fractions or decimals, but percentages are used where appropriate.

[1]Strictly speaking, a percent is not a probability. However, in everyday language, probabilities are often expressed as percents (i.e., there is a 60% chance of rain tomorrow). For this reason, some probabilities will be expressed as percents throughout this book.

Rounding Rule for Probabilities Probabilities should be expressed as reduced fractions or rounded to three decimal places. When the probability of an event is an extremely small decimal, it is permissible to round the decimal to the first nonzero digit after the point. For example, 0.0000587 would be 0.00006. When obtaining probabilities from one of the tables in Appendix A, use the number of decimal places given in the table. If decimals are converted to percentages to express probabilities, move the decimal point two places to the right and add a percent sign.

EXAMPLE 4–5 Drawing Cards

Find the probability of getting a black 6 when one card is randomly selected from an ordinary deck.

SOLUTION

There are 52 cards in an ordinary deck, and there are two black 6s, that is, the 6 of clubs and the 6 of spades. Hence, the probability of getting a black 6 is

$$\frac{2}{52} = \frac{1}{26} \approx 0.038$$

EXAMPLE 4–6 Gender of Children

If a family has three children, find the probability that exactly two of the three children are boys.

SOLUTION

The sample space for the gender of three children has eight outcomes BBB, BBG, BGB, GBB, GGG, GGB, GBG, and BGG. (See Examples 4–3 and 4–4.) Since there are three ways to get two boys and one girl, that is BBG, BGB, and GBB, the probability of having exactly two boys is $\frac{3}{8}$.

Historical Note

Ancient Greeks and Romans made crude dice from animal bones, various stones, minerals, and ivory. When the dice were tested mathematically, some were found to be quite accurate.

In probability theory, it is important to understand the meaning of the words *and* and *or*. For example, if you were asked to find the probability of getting a queen *and* a heart when you were drawing a single card from a deck, you would be looking for the queen of hearts. Here the word *and* means "at the same time." The word *or* has two meanings. For example, if you were asked to find the probability of selecting a queen *or* a heart when one card is selected from a deck, you would be looking for one of the 4 queens or one of the 13 hearts. In this case, the queen of hearts would be included in both cases and counted twice. So there would be $4 + 13 - 1 = 16$ possibilities.

On the other hand, if you were asked to find the probability of getting a queen *or* a king, you would be looking for one of the 4 queens or one of the 4 kings. In this case, there would be $4 + 4 = 8$ possibilities. In the first case, both events can occur at the same time; we say that this is an example of the *inclusive or*. In the second case, both events cannot occur at the same time, and we say that this is an example of the *exclusive or*.

EXAMPLE 4–7 Drawing Cards

A card is drawn from an ordinary deck. Find the probability of getting

a. A heart
b. A black card
c. The 8 of diamonds
d. A queen
e. A face card

SOLUTION

a. Refer to the sample space shown in Figure 4–2. There are 13 hearts in a deck of 52 cards; hence,

$$P(\heartsuit) = \frac{13}{52} = \frac{1}{4} = 0.25$$

b. There are 26 black cards in a deck, that is, 13 clubs and 13 spades. So the probability is

$$P(\text{black card}) = \frac{26}{52} = \frac{1}{2} = 0.5$$

c. There is one 8 of diamonds in a deck of 52 cards, so the probability is

$$P(8\diamondsuit) = \frac{1}{52} \approx 0.019$$

d. There are four queens in a deck of 52 cards; hence,

$$P(\text{queen}) = \frac{4}{52} = \frac{1}{13} \approx 0.077$$

e. There are 12 face cards in an ordinary deck of cards, that is, 4 suits (diamonds, hearts, spades, and clubs) and 3 face cards of each suit (jack, queen, and king), so

$$P(\text{face card}) = \frac{12}{52} = \frac{3}{13} \approx 0.231$$

There are four basic probability rules. These rules are helpful in solving probability problems, in understanding the nature of probability, and in deciding if your answers to the problems are correct.

Historical Note

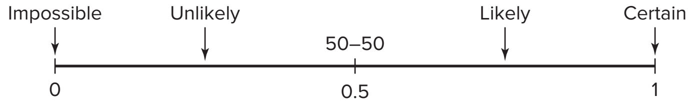
Paintings in tombs excavated in Egypt show that the Egyptians played games of chance. One game called *Hounds and Jackals* played in 1800 B.C. is similar to the present-day game of *Snakes and Ladders*.

Probability Rules

1. The probability of any event E is a number (either a fraction or decimal) between and including 0 and 1. This is denoted by $0 \leq P(E) \leq 1$.
2. The sum of the probabilities of all the outcomes in a sample space is 1.
3. If an event E cannot occur (i.e., the event contains no members in the sample space), its probability is 0.
4. If an event E is certain, then the probability of E is 1.

Rule 1 states that probability values range from 0 to 1. When the probability of an event is close to 0, its occurrence is highly unlikely. When the probability of an event is near 0.5, there is about a 50-50 chance that the event will occur; and when the probability of an event is close to 1, the event is highly likely to occur. See Figure 4–4.

FIGURE 4–4
Range of Probability



Rule 2 can be illustrated by the example of rolling a single die. Each outcome in the sample space has a probability of $\frac{1}{6}$, and the sum of the probabilities of all the outcomes is 1, as shown.

Outcome	1		2		3		4		5		6
Probability	$\frac{1}{6}$		$\frac{1}{6}$		$\frac{1}{6}$		$\frac{1}{6}$		$\frac{1}{6}$		$\frac{1}{6}$
Sum	$\frac{1}{6}$	+	$\frac{1}{6}$	+	$\frac{1}{6}$	+	$\frac{1}{6}$	+	$\frac{1}{6}$	+	$\frac{1}{6} = \frac{6}{6} = 1$

Rule 3 is illustrated in Example 4–8.

EXAMPLE 4–8 Rolling a Die

When a single die is rolled, find the probability of getting a 9.

SOLUTION

Since the sample space is 1, 2, 3, 4, 5, and 6, it is impossible to get a 9. Hence, the probability is $P(9) = \frac{0}{6} = 0$.

Rule 4 states that if $P(E) = 1$, then the event E is certain to occur. This rule is illustrated in Example 4–9.

EXAMPLE 4–9 Rolling a Die

When a single die is rolled, what is the probability of getting a number less than 7?

SOLUTION

Since all outcomes—1, 2, 3, 4, 5, and 6—are less than 7, the probability is

$$P(\text{number less than } 7) = \frac{6}{6} = 1$$

The event of getting a number less than 7 is certain.

Complementary Events

Another important concept in probability theory is that of *complementary events.* When a die is rolled, for instance, the sample space consists of the outcomes 1, 2, 3, 4, 5, and 6. The event E of getting odd numbers consists of the outcomes 1, 3, and 5. The event of not getting an odd number is called the *complement* of event E, and it consists of the outcomes 2, 4, and 6.

The **complement of an event** E is the set of outcomes in the sample space that are not included in the outcomes of event E. The complement of E is denoted by $\overline{E}$ (read "E bar").

Example 4–10 further illustrates the concept of complementary events.

EXAMPLE 4–10 Finding Complements

Find the complement of each event:

a. Selecting a month that has 30 days

b. Selecting a day of the week that begins with the letter S

c. Rolling two dice and getting a number whose sum is 7

d. Selecting a letter of the alphabet (excluding y) that is a vowel

SOLUTION

a. Selecting a month that has 28 or 31 days, that is, January, February, March, May, July, August, October, or December

b. Selecting a day of the week that does not begin with S, that is, Monday, Tuesday, Wednesday, Thursday, or Friday

c. Rolling two dice and getting a sum of 2, 3, 4, 5, 6, 8, 9, 10, 11, or 12

d. Selecting a letter of the alphabet that is a consonant

The outcomes of an event and the outcomes of the complement make up the entire sample space. For example, if two coins are tossed, the sample space is HH, HT, TH, and TT. The complement of "getting all heads" is not "getting all tails," since the event "all heads" is HH, and the complement of HH is HT, TH, and TT. Hence, the complement of the event "all heads" is the event "getting at least one tail."

Since the event and its complement make up the entire sample space, it follows that the sum of the probability of the event and the probability of its complement will equal 1. That is, $P(E) + P(\overline{E}) = 1$. For example, let E = all heads, or HH, and let $\overline{E}$ = at least one tail, or HT, TH, TT. Then $P(E) = \frac{1}{4}$ and $P(\overline{E}) = \frac{3}{4}$; hence, $P(E) + P(\overline{E}) = \frac{1}{4} + \frac{3}{4} = 1$.

The rule for complementary events can be stated algebraically in three ways.

Rule for Complementary Events

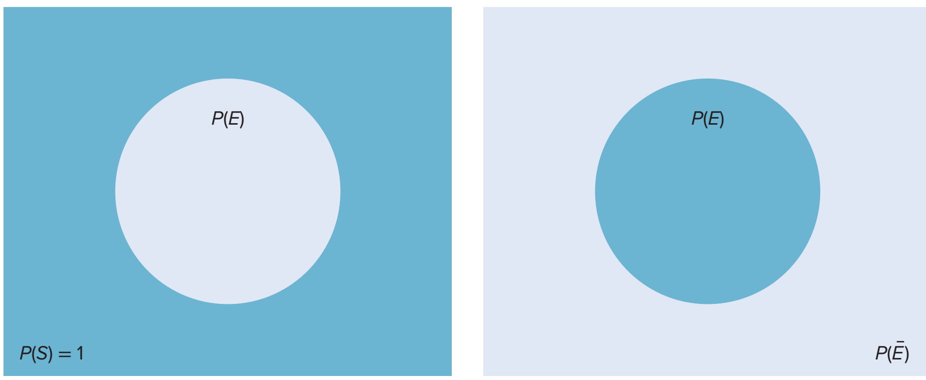
$$P(\overline{E}) = 1 - P(E) \quad \text{or} \quad P(E) = 1 - P(\overline{E}) \quad \text{or} \quad P(E) + P(\overline{E}) = 1$$

Stated in words, the rule is: *If the probability of an event or the probability of its complement is known, then the other can be found by subtracting the probability from* 1. This rule is important in probability theory because at times the best solution to a problem is to find the probability of the complement of an event and then subtract from 1 to get the probability of the event itself.

Probabilities can be represented pictorially by **Venn diagrams.** Figure 4–5(a) shows the probability of a simple event E. The area inside the circle represents the probability of event E, that is, $P(E)$. The area inside the rectangle represents the probability of all the events in the sample space $P(S)$.

The Venn diagram that represents the probability of the complement of an event $P(\overline{E})$ is shown in Figure 4–5(b). In this case, $P(\overline{E}) = 1 - P(E)$, which is the area inside the rectangle but outside the circle representing $P(E)$. Recall that $P(S) = 1$ and $P(E) = 1 - P(\overline{E})$. The reasoning is that $P(E)$ is represented by the area of the circle and $P(\overline{E})$ is the probability of the events that are outside the circle.

FIGURE 4–5
Venn Diagram for the Probability and Complement



(a) Simple probability **(b)** $P(\overline{E}) = 1 - P(E)$

EXAMPLE 4–11 Victims of Violence

In a study, it was found that 24% of people who were victims of a violent crime were age 20 to 24. If a person is selected at random, find the probability that the person is younger than 20 or older than 24.

Source: Based on statistics from the BFI.

SOLUTION

$$\begin{aligned} P(\text{not aged 20 to 24}) &= 1 - P(\text{aged 20 to 24}) \\ &= 1 - 0.24 = 0.76 = 76\% \end{aligned}$$

Empirical Probability

The difference between classical and **empirical probability** is that classical probability assumes that certain outcomes are equally likely (such as the outcomes when a die is rolled), while empirical probability relies on actual experience to determine the likelihood of outcomes. In empirical probability, one might actually roll a given die 6000 times, observe the various frequencies, and use these frequencies to determine the probability of an outcome.

Suppose, for example, that a researcher for the American Automobile Association (AAA) asked 50 people who plan to travel over the Thanksgiving holiday how they will get to their destination. The results can be categorized in a frequency distribution as shown.

Method	Frequency
Drive	41
Fly	6
Train or bus	3
	50

Now probabilities can be computed for various categories. For example, the probability of selecting a person who is driving is $\frac{41}{50}$, since 41 out of the 50 people said that they were driving.

Formula for Empirical Probability

Given a frequency distribution, the probability of an event being in a given class is

$$P(E) = \frac{\text{frequency for the class}}{\text{total frequencies in the distribution}} = \frac{f}{n}$$

This probability is called *empirical probability* and is based on observation.

EXAMPLE 4–12 Travel Survey

In the travel survey just described, find the probability that the person will travel by train or bus over the Thanksgiving holiday.

SOLUTION

$$P(E) = \frac{f}{n} = \frac{3}{50} = 0.06$$

Note: These figures are based on an AAA survey.

EXAMPLE 4–13 Distribution of Blood Types

In a sample of 50 people, 21 had type O blood, 22 had type A blood, 5 had type B blood, and 2 had type AB blood. Set up a frequency distribution and find the following probabilities.

a. A person has type O blood.

b. A person has type A or type B blood.

c. A person has neither type A nor type O blood.

d. A person does not have type AB blood.

Source: The American Red Cross.

SOLUTION

Type	Frequency
A	22
B	5
AB	2
O	21
	Total 50

a. $P(\text{O}) = \dfrac{f}{n} = \dfrac{21}{50}$

b. $P(\text{A or B}) = \dfrac{22}{50} + \dfrac{5}{50} = \dfrac{27}{50}$

(Add the frequencies of the two classes.)

c. $P(\text{neither A nor O}) = \dfrac{5}{50} + \dfrac{2}{50} = \dfrac{7}{50}$

(Neither A nor O means that a person has either type B or type AB blood.)

d. $P(\text{not AB}) = 1 - P(\text{AB}) = 1 - \dfrac{2}{50} = \dfrac{48}{50} = \dfrac{24}{25}$

(Find the probability of not AB by subtracting the probability of type AB from 1.)

EXAMPLE 4–14 Sleep Hours

A recent survey found the following distribution for the number of hours a person sleeps per night.

Less than 6 hours	12
6 hours	26
7 hours	30
8 hours	28
More than 8 hours	4
	100

Find these probabilities for a person selected at random:

a. The person sleeps 8 hours per night.

b. The person sleeps fewer than 7 hours per night.

c. The person sleeps at most 8 hours per night.

d. The person sleeps at least 7 hours per night.

SOLUTION

a. $P(\text{8 hours}) = \dfrac{28}{100} = \dfrac{7}{25}$

b. $P(\text{fewer than 7 hours}) = \dfrac{26}{100} + \dfrac{12}{100} = \dfrac{38}{100} = \dfrac{19}{50}$

c. P(at most 8 hours) = P(At most 8 hours means 8 or less hours)

$= \dfrac{28}{100} + \dfrac{30}{100} + \dfrac{26}{100} + \dfrac{12}{100} = \dfrac{96}{100} = \dfrac{24}{25}$

d. P(at least 7 hours per night) = P(At least 7 hours per night means 7 or more hours per night) $= \dfrac{30}{100} + \dfrac{28}{100} + \dfrac{4}{100} = \dfrac{62}{100} = \dfrac{31}{50}$

Empirical probabilities can also be found by using a relative frequency distribution, as shown in Section 2–2. For example, the relative frequency distribution of the travel survey shown previously is

Method	Frequency	Relative frequency
Drive	41	0.82
Fly	6	0.12
Train or bus	3	0.06
	50	1.00

These frequencies are the same as the relative frequencies explained in Chapter 2.

Law of Large Numbers

When a coin is tossed one time, it is common knowledge that the probability of getting a head is $\frac{1}{2}$. But what happens when the coin is tossed 50 times? Will it come up heads 25 times? Not all the time. You should expect about 25 heads if the coin is fair. But due to chance variation, 25 heads will not occur most of the time.

If the empirical probability of getting a head is computed by using a small number of trials, it is usually not exactly $\frac{1}{2}$. However, as the number of trials increases, the empirical probability of getting a head will approach the theoretical probability of $\frac{1}{2}$, if in fact the coin is fair (i.e., balanced). This phenomenon is an example of the **law of large numbers.**

You should be careful to not think that the number of heads and number of tails tend to "even out." As the number of trials increases, the proportion of heads to the total number of trials will approach $\frac{1}{2}$. This law holds for any type of gambling game—tossing dice, playing roulette, and so on.

It should be pointed out that the probabilities that the proportions steadily approach may or may not agree with those theorized in the classical model. If not, it can have important implications, such as "the die is not fair." Pit bosses in Las Vegas watch for empirical trends that do not agree with classical theories, and they will sometimes take a set of dice out of play if observed frequencies are too far out of line with classical expected frequencies.

Subjective Probability

The third type of probability is called *subjective probability.* **Subjective probability** uses a probability value based on an educated guess or estimate, employing opinions and inexact information.

In subjective probability, a person or group makes an educated guess at the chance that an event will occur. This guess is based on the person's experience and evaluation of a solution. For example, a sportswriter may say that there is a 70% probability that the Pirates will win the pennant next year. A physician might say that, on the basis of her diagnosis, there is a 30% chance the patient will need an operation. A seismologist might say there is an 80% probability that an earthquake will occur in a certain area. These are only a few examples of how subjective probability is used in everyday life.

All three types of probability (classical, empirical, and subjective) are used to solve a variety of problems in business, engineering, and other fields.

Probability and Risk Taking

An area in which people fail to understand probability is risk taking. Actually, people fear situations or events that have a relatively small probability of happening rather than those events that have a greater likelihood of occurring. For example, many people think that the crime rate is increasing every year. However, in his book entitled *How Risk Affects Your Everyday Life,* author James Walsh states: "Despite widespread concern about the number of crimes committed in the United States, FBI and Justice Department statistics show that the national crime rate has remained fairly level for 20 years. It even dropped slightly in the early 1990s."

He further states, "Today most media coverage of risk to health and well-being focuses on shock and outrage." Shock and outrage make good stories and can scare us about the wrong dangers. For example, the author states that if a person is 20% overweight, the loss of life expectancy is 900 days (about 3 years), but loss of life expectancy from exposure to radiation emitted by nuclear power plants is 0.02 day. As you can see, being overweight is much more of a threat than being exposed to radioactive emission.

Many people gamble daily with their lives, for example, by using tobacco, drinking and driving, and riding motorcycles. When people are asked to estimate the probabilities or frequencies of death from various causes, they tend to overestimate causes such as accidents, fires, and floods and to underestimate the probabilities of death from diseases (other than cancer), strokes, etc. For example, most people think that their chances of dying of a heart attack are 1 in 20, when in fact they are almost 1 in 3; the chances of dying by pesticide poisoning are 1 in 200,000 (*True Odds* by James Walsh). The reason people think this way is that the news media sensationalize deaths resulting from catastrophic events and rarely mention deaths from disease.

When you are dealing with life-threatening catastrophes such as hurricanes, floods, automobile accidents, or texting while driving, it is important to get the facts. That is, get the actual numbers from accredited statistical agencies or reliable statistical studies, and then compute the probabilities and make decisions based on your knowledge of probability and statistics.

In summary, then, when you make a decision or plan a course of action based on probability, make sure that you understand the true probability of the event occurring. Also, find out how the information was obtained (i.e., from a reliable source). Weigh the cost of the action and decide if it is worth it. Finally, look for other alternatives or courses of action with less risk involved.

Applying the Concepts **4–1**

Tossing a Coin

Assume you are at a carnival and decide to play one of the games. You spot a table where a person is flipping a coin, and since you have an understanding of basic probability, you believe that the odds of winning are in your favor. When you get to the table, you find out that all you have to do is to guess which side of the coin will be facing up after it is tossed. You are assured that the coin is fair, meaning that each of the two sides has an equally likely chance of occurring. You think back about what you learned in your statistics class about probability before you decide what to bet on. Answer the following questions about the coin-tossing game.

1. What is the sample space?
2. What are the possible outcomes?
3. What does the classical approach to probability say about computing probabilities for this type of problem?

You decide to bet on heads, believing that it has a 50% chance of coming up. A friend of yours, who had been playing the game for awhile before you got there, tells you that heads has come up the last 9 times in a row. You remember the law of large numbers.

4. What is the law of large numbers, and does it change your thoughts about what will occur on the next toss?
5. What does the empirical approach to probability say about this problem, and could you use it to solve this problem?
6. Can subjective probabilities be used to help solve this problem? Explain.
7. Assume you could win $1 million if you could guess what the results of the next toss will be. What would you bet on? Why?

See pages 253–255 for the answers.

Exercises 4–1

1. What is a probability experiment?

2. Define *sample space*.

3. What is the difference between an outcome and an event?

4. What are equally likely events?

5. What is the range of the values of the probability of an event?

6. When an event is certain to occur, what is its probability?

7. If an event cannot happen, what value is assigned to its probability?

8. What is the sum of the probabilities of all the outcomes in a sample space?

9. If the probability that it will rain tomorrow is 0.20, what is the probability that it won't rain tomorrow? Would you recommend taking an umbrella?

10. A probability experiment is conducted. Which of these cannot be considered a probability outcome?

a. $\frac{2}{3}$ *d.* 1.65 *g.* 1

b. 0.63 *e.* -0.44 *h.* 125%

c. $-\frac{3}{5}$ *f.* 0 *i.* 24%

11. Classify each statement as an example of classical probability, empirical probability, or subjective probability.

a. The probability that a person will watch the 6 o'clock evening news is 0.15.

b. The probability of winning at a Chuck-a-Luck game is $\frac{5}{36}$.

c. The probability that a bus will be in an accident on a specific run is about 6%.

d. The probability of getting a royal flush when five cards are selected at random is $\frac{1}{649,740}$.

12. Classify each statement as an example of classical probability, empirical probability, or subjective probability.

a. The probability that a student will get a C or better in a statistics course is about 70%.

b. The probability that a new fast-food restaurant will be a success in Chicago is 35%.

c. The probability that interest rates will rise in the next 6 months is 0.50.

d. The probability that the unemployment rate will fall next month is 0.03.

13. Rolling a Die If a die is rolled one time, find these probabilities:

a. Getting a 7

b. Getting an odd number

c. Getting a number less than 7

d. Getting a prime number (2, 3, or 5)

14. Rolling a Die If a die is rolled one time, find these probabilities:

a. Getting a number less than 7.

b. Getting a number greater than or equal to 3

c. Getting a number greater than 2 and an even number

d. Getting a number less than 1

15. Rolling Two Dice If two dice are rolled one time, find the probability of getting these results:

a. A sum of 5

b. A sum of 9 or 10

c. Doubles

16. Rolling Two Dice If two dice are rolled one time, find the probability of getting these results:

a. A sum less than 9

b. A sum greater than or equal to 10

c. A 3 on one die or on both dice.

17. Drawing a Card If one card is drawn from a deck, find the probability of getting these results:

a. An ace

b. A heart

c. A 6 of spades

d. A 10 or a jack

e. A card whose face values less than 7 (Count aces as 1.)

18. Drawing a Card If a card is drawn from a deck, find the probability of getting these results:

a. A 6 and a spade

b. A black king

c. A red card and a 7

d. A diamond or a heart

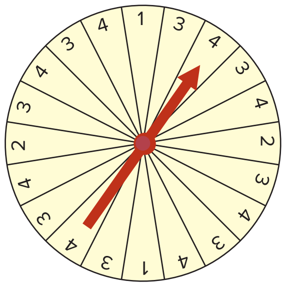
e. A black card

19. Shopping Mall Promotion A shopping mall has set up a promotion as follows. With any mall purchase of \$50 or more, the customer gets to spin the wheel shown here. If a number 1 comes up, the customer wins \$10. If the number 2 comes up, the customer wins \$5; and if the number 3 or 4 comes up, the customer wins a discount coupon. Find the following probabilities.

a. The customer wins \$10.

b. The customer wins money.

c. The customer wins a coupon.



20. Selecting a State Choose one of the 50 states at random.

a. What is the probability that it begins with the letter M?

b. What is the probability that it doesn't begin with a vowel?

21. Human Blood Types Human blood is grouped into four types. The percentages of Americans with each type are listed below.

O 43%	A 40%	B 12%	AB 5%

Choose one American at random. Find the probability that this person

a. Has type B blood

b. Has type AB or O blood

c. Does not have type O blood

22. 2014 Top Albums (Based on U.S. sales) Of all of the U.S. album sales 1989 (Taylor Swift) accounted for 25% of sales, Frozen (Various Artists) accounted for 24.1% of sales, In the Lonely Hour (Sam Smith) accounted for 8.2% of sales. What is the probability that a randomly selected album was something other than these three albums?

Source: 2014 Nielsen Music U.S. Report.

23. Prime Numbers A prime number is a number that is evenly divisible only by 1 and itself. The prime numbers less than 100 are listed below.

2	3	5	7	11	13	17	19	23	29	31
37	41	43	47	53	59	61	67	71	73	79
83	89	97								

Choose one of these numbers at random. Find the probability that

a. The number is odd

b. The sum of the digits is odd

c. The number is greater than 70

24. Rural Speed Limits Rural speed limits for all 50 states are indicated below.

60 mph	65 mph	70 mph	75 mph
1 (HI)	18	18	13

Choose one state at random. Find the probability that its speed limit is

a. 60 or 70 miles per hour

b. Greater than 65 miles per hour

c. 70 miles per hour or less

Source: World Almanac.

25. Gender of Children A couple has 4 children. Find each probability.

a. All girls

b. Exactly two girls and two boys

c. At least one child who is a girl

d. At least one child of each gender

26. Sources of Energy Uses in the United States A breakdown of the sources of energy used in the United States is shown below. Choose one energy source at random. Find the probability that it is

a. Not oil

b. Natural gas or oil

c. Nuclear

Oil 39%	Natural gas 24%	Coal 23%
Nuclear 8%	Hydropower 3%	Other 3%

Source: www.infoplease.com

27. Game of Craps In a game of craps, a player wins on the first roll if the player rolls a sum of 7 or 11, and the player loses if the player rolls a 2, 3, or 12. Find the probability that the game will last only one roll.

28. Computers in Elementary Schools Elementary and secondary schools were classified by the number of computers they had.

Computers	1–10	11–20	21–50	51–100	100+
Schools	3170	4590	16,741	23,753	34,803

Choose one school at random. Find the probability that it has

a. 50 or fewer computers

b. More than 100 computers

c. No more than 20 computers

Source: World Almanac.

29. College Debt The following information shows the amount of debt students who graduated from college incur for a specific year.

\$1 to \$5000	**\$5001 to \$20,000**	**\$20,001 to \$50,000**	**\$50,000+**
27%	40%	19%	14%

If a person who graduates has some debt, find the probability that

a. It is less than \$5001

b. It is more than \$20,000

c. It is between \$1 and \$20,000

d. It is more than \$50,000

Source: USA TODAY.

30. Population of Hawaii The population of Hawaii is 22.7% white, 1.5% African-American, 37.7% Asian, 0.2% Native American/Alaskan, 9.46% Native Hawaiian/Pacific Islander, 8.9% Hispanic, 19.4% two or more races, and 0.14% some other. Choose one Hawaiian resident at random. What is the probability that he/she is a Native Hawaiian or Pacific Islander? Asian? White?

31. Crimes Committed The numbers show the number of crimes committed in a large city. If a crime is selected at random, find the probability that it is a motor vehicle theft. What is the probability that it is not an assault?

Theft	1375
Burglary of home or office	500
Motor vehicle theft	275
Assault	200
Robbery	125
Rape or homicide	25

Source: Based on FBI statistics.

32. Living Arrangements for Children Here are the living arrangements of children under 18 years old living in the United States in a recent year. Numbers are in thousands.

Both parents	51,823
Mother only	17,283
Father only	2,572
Neither parent	3,041

Choose one child at random; what is the probability that the child lives with both parents? With the mother present?

Source: Time Almanac.

33. Motor Vehicle Accidents During a recent year, there were 13.5 million automobile accidents, 5.2 million truck accidents, and 178,000 motorcycle accidents. If one accident is selected at random, find the probability that it is either a truck or motorcycle accident. What is the probability that it is not a truck accident?

Source: Based on data from the National Safety Council.

34. Federal Government Revenue The source of federal government revenue for a specific year is

50% from individual income taxes
32% from social insurance payroll taxes
10% from corporate income taxes
3% from excise taxes
5% other

If a revenue source is selected at random, what is the probability that it comes from individual or corporate income taxes?

Source: New York Times Almanac.

35. Selecting a Bill A box contains a $1 bill, a $5 bill, a $10 bill, and a $20 bill. A bill is selected at random, and it is not replaced; then a second bill is selected at random. Draw a tree diagram and determine the sample space.

36. Tossing Coins Draw a tree diagram and determine the sample space for tossing four coins.

37. Selecting Numbered Balls Four balls numbered 1 through 4 are placed in a box. A ball is selected at random, and its number is noted; then it is replaced. A second ball is selected at random, and its number is noted. Draw a tree diagram and determine the sample space.

38. Family Dinner Combinations A family special at a neighborhood restaurant offers dinner for four for $39.99. There are 3 appetizers available, 4 entrees, and 3 desserts from which to choose. The special includes one of each. Represent the possible dinner combinations with a tree diagram.

39. Required First-Year College Courses First-year students at a particular college must take one English class, one class in mathematics, a first-year seminar, and an elective. There are 2 English classes to choose from, 3 mathematics classes, 5 electives, and everyone takes the same first-year seminar. Represent the possible schedules, using a tree diagram.

40. Tossing a Coin and Rolling a Die A coin is tossed; if it falls heads up, it is tossed again. If it falls tails up, a die is rolled. Draw a tree diagram and determine the outcomes.

Extending the Concepts

41. Distribution of CEO Ages The distribution of ages of CEOs is as follows:

Age	Frequency
21–30	1
31–40	8
41–50	27
51–60	29
61–70	24
71–up	11

Source: Information based on *USA TODAY* Snapshot.

If a CEO is selected at random, find the probability that his or her age is

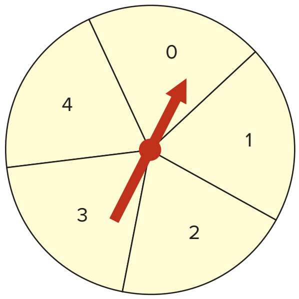
a. Between 31 and 40
b. Under 31
c. Over 30 and under 51
d. Under 31 or over 60

42. Tossing a Coin A person flipped a coin 100 times and obtained 73 heads. Can the person conclude that the coin was unbalanced?

43. Medical Treatment A medical doctor stated that with a certain treatment, a patient has a 50% chance of

recovering without surgery. That is, "Either he will get well or he won't get well." Comment on this statement.

44. Wheel Spinner The wheel spinner shown here is spun twice. Find the sample space, and then determine the probability of the following events.



a. An odd number on the first spin and an even number on the second spin (*Note:* 0 is considered even.)

b. A sum greater than 4

c. Even numbers on both spins

d. A sum that is odd

e. The same number on both spins

45. Tossing Coins Toss three coins 128 times and record the number of heads (0, 1, 2, or 3); then record your results with the theoretical probabilities. Compute the empirical probabilities of each.

46. Tossing Coins Toss two coins 100 times and record the number of heads (0, 1, 2). Compute the probabilities of each outcome, and compare these probabilities with the theoretical results.

47. Odds Odds are used in gambling games to make them fair. For example, if you rolled a die and won every time you rolled a 6, then you would win on average once every 6 times. So that the game is fair, the odds of 5 to 1 are given. This means that if you bet $1 and won, you could win $5. On average, you would win $5 once in 6 rolls and lose $1 on the other 5 rolls—hence the term *fair game.*

In most gambling games, the odds given are not fair. For example, if the odds of winning are really 20 to 1, the house might offer 15 to 1 in order to make a profit.

Odds can be expressed as a fraction or as a ratio, such as $\frac{5}{1}$, 5:1, or 5 to 1. Odds are computed in favor of the event or against the event. The formulas for odds are

$$\text{Odds in favor} = \frac{P(E)}{1 - P(E)}$$

$$\text{Odds against} = \frac{P(\overline{E})}{1 - P(\overline{E})}$$

In the die example,

$$\text{Odds in favor of a 6} = \frac{\frac{1}{6}}{\frac{5}{6}} = \frac{1}{5} \text{ or } 1{:}5$$

$$\text{Odds against a 6} = \frac{\frac{5}{6}}{\frac{1}{6}} = \frac{5}{1} \text{ or } 5{:}1$$

Find the odds in favor of and against each event.

a. Rolling a die and getting a 2

b. Rolling a die and getting an even number

c. Drawing a card from a deck and getting a spade

d. Drawing a card and getting a red card

e. Drawing a card and getting a queen

f. Tossing two coins and getting two tails

g. Tossing two coins and getting exactly one tail

4–2 The Addition Rules for Probability

OBJECTIVE 2

Find the probability of compound events, using the addition rules.

Many problems involve finding the probability of two or more events. For example, at a large political gathering, you might wish to know, for a person selected at random, the probability that the person is a female or is a Republican. In this case, there are three possibilities to consider:

1. The person is a female.
2. The person is a Republican.
3. The person is both a female and a Republican.

Consider another example. At the same gathering there are Republicans, Democrats, and Independents. If a person is selected at random, what is the probability that the person is a Democrat or an Independent? In this case, there are only two possibilities:

1. The person is a Democrat.
2. The person is an Independent.

The difference between the two examples is that in the first case, the person selected can be a female and a Republican at the same time. In the second case, the person selected cannot be both a Democrat and an Independent at the same time. In the second case,

Historical Note

The first book on probability, *The Book of Chance and Games,* was written by Jerome Cardan (1501–1576). Cardan was an astrologer, philosopher, physician, mathematician, and gambler. This book contained techniques on how to cheat and how to catch others at cheating.

the two events are said to be *mutually exclusive;* in the first case, they are not mutually exclusive.

Two events are **mutually exclusive events** or **disjoint events** if they cannot occur at the same time (i.e., they have no outcomes in common).

In another situation, the events of getting a 4 and getting a 6 when a single card is drawn from a deck are mutually exclusive events, since a single card cannot be both a 4 and a 6. On the other hand, the events of getting a 4 and getting a heart on a single draw are not mutually exclusive, since you can select the 4 of hearts when drawing a single card from an ordinary deck.

EXAMPLE 4–15 Determining Mutually Exclusive Events

Determine whether the two events are mutually exclusive. Explain your answer.

a. Randomly selecting a female student
Randomly selecting a student who is a junior

b. Randomly selecting a person with type A blood
Randomly selecting a person with type O blood

c. Rolling a die and getting an odd number
Rolling a die and getting a number less than 3

d. Randomly selecting a person who is under 21 years of age
Randomly selecting a person who is over 30 years of age

SOLUTION

a. These events are not mutually exclusive since a student can be both female and a junior.

b. These events are mutually exclusive since a person cannot have type A blood and type O blood at the same time.

c. These events are not mutually exclusive since the number 1 is both an odd number and a number less than 3.

d. These events are mutually exclusive since a person cannot be both under 21 and over 30 years of age at the same time.

EXAMPLE 4–16 Drawing a Card

Determine which events are mutually exclusive and which events are not mutually exclusive when a single card is drawn at random from a deck.

a. Getting a face card; getting a 6

b. Getting a face card; getting a heart

c. Getting a 7; getting a king

d. Getting a queen; getting a spade

SOLUTION

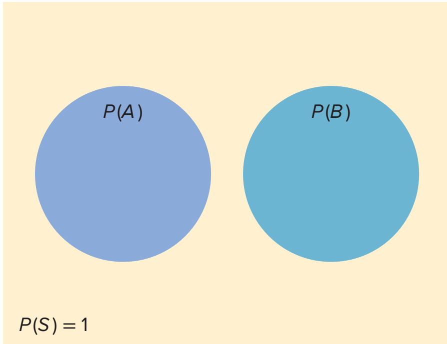
a. These events are mutually exclusive since you cannot draw one card that is both a face card (jack, queen, or king) and a card that is a 6 at the same time.

b. These events are not mutually exclusive since you can get one card that is a face card and is a heart: that is, a jack of hearts, a queen of hearts, or a king of hearts.

c. These events are mutually exclusive since you cannot get a single card that is both a 7 and a king.

d. These events are not mutually exclusive since you can get the queen of spades.

FIGURE 4–6
Venn Diagram for Addition Rule 1 When the Events Are Mutually Exclusive



Mutually exclusive events
$P(A \text{ or } B) = P(A) + P(B)$

The probability of two or more events can be determined by the *addition rules.* The first addition rule is used when the events are mutually exclusive.

Addition Rule 1

When two events A and B are mutually exclusive, the probability that A or B will occur is

$$P(A \text{ or } B) = P(A) + P(B)$$

Figure 4–6 shows a Venn diagram that represents two mutually exclusive events A and B. In this case, $P(A \text{ or } B) = P(A) + P(B)$, since these events are mutually exclusive and do not overlap. In other words, the probability of occurrence of event A or event B is the sum of the areas of the two circles.

EXAMPLE 4–17 Endangered Species

In the United States there are 59 different species of mammals that are endangered, 75 different species of birds that are endangered, and 68 species of fish that are endangered. If one animal is selected at random, find the probability that it is either a mammal or a fish.

Source: Based on information from the U.S. Fish and Wildlife Service.

SOLUTION

Since there are 59 species of mammals and 68 species of fish that are endangered and a total of 202 endangered species, $P(\text{mammal or fish}) = P(\text{mammal}) + P(\text{fish}) = \frac{59}{202} + \frac{68}{202} = \frac{127}{202} = 0.629$. The events are mutually exclusive.

EXAMPLE 4–18 Research and Development Employees

The corporate research and development centers for three local companies have the following numbers of employees:

U.S. Steel	110
Alcoa	750
Bayer Material Science	250

If a research employee is selected at random, find the probability that the employee is employed by U.S. Steel or Alcoa.

Source: Pittsburgh Tribune Review.

SOLUTION

$$P(\text{U.S. Steel or Alcoa}) = P(\text{U.S. Steel}) + P(\text{Alcoa})$$
$$= \frac{110}{1110} + \frac{750}{1110} = \frac{860}{1110} = \frac{86}{111} = 0.775$$

EXAMPLE 4–19 Favorite Ice Cream

In a survey, 8% of the respondents said that their favorite ice cream flavor is cookies and cream, and 6% like mint chocolate chip. If a person is selected at random, find the probability that her or his favorite ice cream flavor is either cookies and cream or mint chocolate chip.

Source: Rasmussen Report.

SOLUTION

$$\begin{aligned} &P(\text{cookies and cream or mint chocolate chip}) \\ &\quad = P(\text{cookies and cream}) + P(\text{mint chocolate chip}) \\ &\quad = 0.08 + 0.06 = 0.14 = 14\% \end{aligned}$$

These events are mutually exclusive.

Historical Note

Venn diagrams were developed by mathematician John Venn (1834–1923) and are used in set theory and symbolic logic. They have been adapted to probability theory also. In set theory, the symbol $\cup$ represents the *union* of two sets, and $A \cup B$ corresponds to A or B. The symbol $\cap$ represents the *intersection* of two sets, and $A \cap B$ corresponds to A and B. Venn diagrams show only a general picture of the probability rules and do not portray all situations, such as $P(A) = 0$, accurately.

The probability rules can be extended to three or more events. For three mutually exclusive events A, B, and C,

$$P(A \text{ or } B \text{ or } C) = P(A) + P(B) + P(C)$$

When events are not mutually exclusive, addition rule 2 can be used to find the probability of the events.

Addition Rule 2

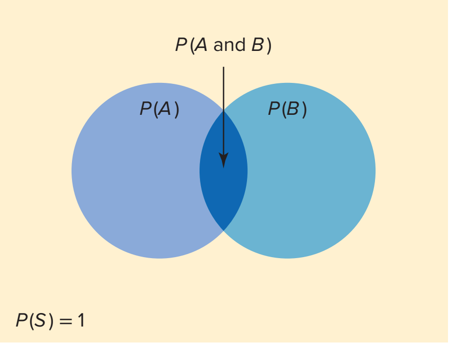
If A and B are *not* mutually exclusive, then

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

Note: This rule can also be used when the events are mutually exclusive, since $P(A \text{ and } B)$ will always equal 0. However, it is important to make a distinction between the two situations.

Figure 4–7 represents the probability of two events that are *not* mutually exclusive. In this case, $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$. The area in the intersection or overlapping part of both circles corresponds to $P(A \text{ and } B)$; and when the area of circle A is added to the area of circle B, the overlapping part is counted twice. It must therefore be subtracted once to get the correct area or probability.

FIGURE 4–7
Venn Diagram for Addition Rule 2 When Events Are Not Mutually Exclusive



Nonmutually exclusive events
$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

EXAMPLE 4–20 Drawing a Card

A single card is drawn at random from an ordinary deck of cards. Find the probability that it is either a 9 or a diamond.

SOLUTION

There are 4 nines and 13 diamonds in a deck of cards, and one of the 9s is a diamond, so it is counted twice. Hence,

$$P(9 \text{ or } \blacklozenge) = P(9) + P(\blacklozenge) - P(9\blacklozenge) = \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52} = \frac{4}{13} \approx 0.308$$

Interesting Fact

Card Shuffling

How many times does a deck of cards need to be shuffled so that the cards are in random order? Actually, this question is not easy to answer since there are many variables. First, several different methods are used to shuffle a deck of cards. Some of the methods are the riffle method, the overhand method, the Corgi method, and the Faro method.

Another factor that needs to be considered is what is meant by the cards being in a random order. There are several statistical tests that can be used to determine if a deck of cards is randomized after several shuffles, but these tests give somewhat different results.

Two mathematicians, Persi Diaconis and Dave Bayer, concluded that a deck of cards starts to become random after 5 good shuffles and is completely random after 7 shuffles. However, a later study done by Trefthen concluded that only 6 shuffles are necessary. The difference was based on what is considered a randomized deck of cards.

EXAMPLE 4–21 Selecting a Medical Staff Person

In a hospital unit there are 8 nurses and 5 physicians; 7 nurses and 3 physicians are females. If a staff person is selected, find the probability that the subject is a nurse or a male.

SOLUTION

The sample space is shown here.

Staff	Females	Males	Total
Nurses	7	1	8
Physicians	3	2	5
Total	10	3	13

The probability is

$$P(\text{nurse or male}) = P(\text{nurse}) + P(\text{male}) - P(\text{male nurse})$$
$$= \frac{8}{13} + \frac{3}{13} - \frac{1}{13} = \frac{10}{13} \approx 0.769$$

EXAMPLE 4–22 Driving While Intoxicated

On New Year's Eve, the probability of a person driving while intoxicated is 0.32, the probability of a person having a driving accident is 0.09, and the probability of a person having a driving accident while intoxicated is 0.06. What is the probability of a person driving while intoxicated or having a driving accident?

SOLUTION

$$\begin{aligned} P(\text{intoxicated or accident}) &= P(\text{intoxicated}) + P(\text{accident}) \\ &\quad - P(\text{intoxicated and accident}) \\ &= 0.32 + 0.09 - 0.06 = 0.35 \end{aligned}$$

For three events that are *not* mutually exclusive,

$$\begin{aligned} P(A \text{ or } B \text{ or } C) = P(A) + P(B) + P(C) - P(A \text{ and } B) - P(A \text{ and } C) \\ - P(B \text{ and } C) + P(A \text{ and } B \text{ and } C) \end{aligned}$$

See Exercises 23 and 24 in this section.

In summary, then, when the two events are mutually exclusive, use addition rule 1. When the events are not mutually exclusive, use addition rule 2.

Applying the Concepts 4–2

Which Pain Reliever Is Best?

Assume that following an injury you received from playing your favorite sport, you obtain and read information on new pain medications. In that information you read of a study that was conducted to test the side effects of two new pain medications. Use the following table to answer the questions and decide which, if any, of the two new pain medications you will use.

	Number of side effects in 12-week clinical trial		
Side effect	**Placebo** $n = 192$	**Drug A** $n = 186$	**Drug B** $n = 188$
Upper respiratory congestion	10	32	19
Sinus headache	11	25	32
Stomachache	2	46	12
Neurological headache	34	55	72
Cough	22	18	31
Lower respiratory congestion	2	5	1

1. How many subjects were in the study?
2. How long was the study?
3. What were the variables under study?
4. What type of variables are they, and what level of measurement are they on?
5. Are the numbers in the table exact figures?
6. What is the probability that a randomly selected person was receiving a placebo?
7. What is the probability that a person was receiving a placebo or drug A? Are these mutually exclusive events? What is the complement to this event?
8. What is the probability that a randomly selected person was receiving a placebo or experienced a neurological headache?
9. What is the probability that a randomly selected person was not receiving a placebo or experienced a sinus headache?

See page 254 for the answers.

Exercises 4–2

1. Define mutually exclusive events, and give an example of two events that are mutually exclusive and two events that are not mutually exclusive.

2. Explain briefly why addition rule 2 can be used when two events are mutually exclusive.

3. Cards, Dice, and Students Determine whether these events are mutually exclusive:

a. Draw a card: get a spade and get a 6

b. Roll a die: get a prime number (2, 3, 5)

c. Roll two dice: get a sum of 7 or get a sum that is an even number

d. Select a student at random in your class: get a male or get a sophomore

4. Determine whether these events are mutually exclusive.

a. Roll two dice: Get a sum of 7 or get doubles.

b. Select a student in your college: The student is a sophomore and the student is a business major.

c. Select any course: It is a calculus course and it is an English course.

d. Select a registered voter: The voter is a Republican and the voter is a Democrat.

5. College Degrees Awarded The table below represents the college degrees awarded in a recent academic year by gender.

	Bachelor's	Master's	Doctorate
Men	573,079	211,381	24,341
Women	775,424	301,264	21,683

Choose a degree at random. Find the probability that it is

a. A bachelor's degree

b. A doctorate or a degree awarded to a woman

c. A doctorate awarded to a woman

d. Not a master's degree

Source: www.nces.ed.gov

6. **Riding to School** The probability that John will drive to school is 0.37, the probability that he will ride with friends is 0.23, and the probability that his parents will take him is 0.4. He is not allowed to have passengers in the car when he is driving. What is the probability that John will have company on the way to school?

7. **Medical Specialities** The following doctors were observed on staff at a local hospital.

	MD	Doctor of Osteopathy
Pathology	6	1
Pediatrics	7	2
Orthopedics	20	2

Choose one doctor at random; what is the probability that

a. She is a pathologist?

b. He is an orthopedist or an MD?

8. **U.S. Population** The data show the U.S. population by age.

Under 20 years	27.0%
20 years and over	73.0
65 years and over	13.1

Choose one person from the United States at random. Find the probability that the person is

a. From 20 years to 64 years

b. Under 20 or 65 and over

c. Not 65 and over

9. **Snack Foods** In a meeting room in a dormitory there are 8 bags of potato chips, 5 bags of popcorn, 2 bags of pretzels, and 1 bag of cheese puffs. If a student selects 1 bag at random, find the probability that it is a bag of potato chips or a bag of pretzels.

10. **Selecting a Movie** A media rental store rented the following number of movie titles in each of these categories: 170 horror, 230 drama, 120 mystery, 310 romance, and 150 comedy. If a person selects a movie to rent, find the probability that it is a romance or a comedy. Is this event likely or unlikely to occur? Explain your answer.

11. **Pizza Sales** A pizza restaurant sold 24 cheese pizzas and 16 pizzas with one or more toppings. Twelve of the cheese pizzas were eaten at work, and 10 of the pizzas with one or more toppings were eaten at work. If a pizza was selected at random, find the probability of each:

a. It was a cheese pizza eaten at work.

b. It was a pizza with either one or more toppings, and it was not eaten at work.

c. It was a cheese pizza, or it was a pizza eaten at work.

12. **Selecting a Book** At a used-book sale, 100 books are adult books and 160 are children's books. Of the adult books, 70 are nonfiction while 60 of the children's books are nonfiction. If a book is selected at random, find the probability that it is

a. Fiction

b. Not a children's nonfiction book

c. An adult book or a children's nonfiction book

13. **Young Adult Residences** According to the Bureau of the Census, the following statistics describe the number (in thousands) of young adults living at home or in a dormitory in the year 2004.

	Ages 18–24	Ages 25–34
Male	7922	2534
Female	5779	995

Source: World Almanac.

Choose one student at random. Find the probability that the student is

a. A female student aged 25–34 years

b. Male or aged 18–24 years

c. Under 25 years of age and not male

14. **Endangered Species** The chart below shows the numbers of endangered and threatened species both here in the United States and abroad.

	Endangered		Threatened	
	United States	Foreign	United States	Foreign
Mammals	68	251	10	20
Birds	77	175	13	6
Reptiles	14	64	22	16
Amphibians	11	8	10	1

Source: www.infoplease.com

Choose one species at random. Find the probability that it is

a. Threatened and in the United States

b. An endangered foreign bird

c. A mammal or a threatened foreign species

15. **Multiple Births** The number of multiple births in the United States for a recent year indicated that there were 128,665 sets of twins, 7110 sets of triplets, 468 sets of quadruplets, and 85 sets of quintuplets. Choose one set of siblings at random.

a. Find the probability that it represented more than two babies.

b. Find the probability that it represented quads or quints.

c. Now choose one baby from these multiple births. What is the probability that the baby was a triplet?

16. Licensed Drivers in the United States In a recent year there were the following numbers (in thousands) of licensed drivers in the United States.

	Male	Female
Age 19 and under	4746	4517
Age 20	1625	1553
Age 21	1679	1627

Source: World Almanac.

Choose one driver at random. Find the probability that the driver is

a. Male and 19 years or under
b. Age 20 or female
c. At least 20 years old

17. Prison Education In a federal prison, inmates can select to complete high school, take college courses, or do neither. The following survey results were obtained using ages of the inmates.

Age	High School Courses	College Courses	Neither
Under 30	53	107	450
30 and over	27	32	367

If a prisoner is selected at random, find these probabilities:

a. The prisoner does not take classes.
b. The prisoner is under 30 and is taking either a high school class or a college class.
c. The prisoner is over 30 and is taking either a high school class or a college class.

18. Mail Delivery A local postal carrier distributes first-class letters, advertisements, and magazines. For a certain day, she distributed the following numbers of each type of item.

Delivered to	First-class letters	Ads	Magazines
Home	325	406	203
Business	732	1021	97

If an item of mail is selected at random, find these probabilities.

a. The item went to a home.
b. The item was an ad, or it went to a business.
c. The item was a first-class letter, or it went to a home.

19. Medical Tests on Emergency Patients The frequency distribution shown here illustrates the number of medical tests conducted on 30 randomly selected emergency patients.

Number of tests performed	Number of patients
0	12
1	8
2	2
3	3
4 or more	5

If a patient is selected at random, find these probabilities.

a. The patient has had exactly 2 tests done.
b. The patient has had at least 2 tests done.
c. The patient has had at most 3 tests done.
d. The patient has had 3 or fewer tests done.
e. The patient has had 1 or 2 tests done.

20. College Fundraiser A social organization of 32 members sold college sweatshirts as a fundraiser. The results of their sale are shown below.

No. of sweatshirts	No. of students
0	2
1–5	13
6–10	8
11–15	4
16–20	4
20+	1

Choose one student at random. Find the probability that the student sold

a. More than 10 sweatshirts
b. At least one sweatshirt
c. 1–5 or more than 15 sweatshirts

21. Emergency Room Tests The frequency distribution shows the number of medical tests conducted on 30 randomly selected emergency room patients.

Number of tests performed	Number of patients
0	11
1	9
2	5
3	4
4 or more	1

If a patient is selected at random, find these probabilities:

a. The patient had exactly 3 tests done.
b. The patient had at most 2 tests done.
c. The patient has 1 or 2 tests done.
d. The patient had fewer than 3 tests done.
e. The patient had at least 3 tests done.

22. Medical Patients A recent study of 300 patients found that of 100 alcoholic patients, 87 had elevated cholesterol levels, and of 200 nonalcoholic patients, 43 had elevated cholesterol levels. If a patient is selected at random, find the probability that the patient is the following.

a. An alcoholic with elevated cholesterol level
b. A nonalcoholic
c. A nonalcoholic with nonelevated cholesterol level

23. Selecting a Card If one card is drawn from an ordinary deck of cards, find the probability of getting each event:

a. A 7 or an 8 or a 9
b. A spade or a queen or a king
c. A club or a face card

d. An ace or a diamond or a heart

e. A 9 or a 10 or a spade or a club

24. Rolling Die Two dice are rolled. Find the probability of getting

a. A sum of 8, 9, or 10

b. Doubles or a sum of 7

c. A sum greater than 9 or less than 4

d. Based on the answers to *a*, *b*, and *c*, which is least likely to occur?

25. Apple Production For a recent year, about 11 billion pounds of apples were harvested. About 4.4 billion pounds of apples were made into apple juice, about 1 billion pounds of apples were made into apple sauce, and 1 billion pounds of apples were used for other commercial purposes. If 1 billion pounds of apples were selected at random, what is the probability that the apples were used for apple juice or applesauce?

Source: International Apple Institute.

26. Rolling Dice Three dice are rolled. Find the probability of getting

a. Triples *b.* A sum of 5

Extending the Concepts

27. Purchasing a Pizza The probability that a customer selects a pizza with mushrooms or pepperoni is 0.55, and the probability that the customer selects only mushrooms is 0.32. If the probability that he or she selects only pepperoni is 0.17, find the probability of the customer selecting both items.

28. Building a New Home In building new homes, a contractor finds that the probability of a home buyer selecting a two-car garage is 0.70 and of selecting a one-car garage is 0.20. Find the probability that the buyer will select no garage. The builder does not build houses with three-car or more garages.

29. In Exercise 28, find the probability that the buyer will not want a two-car garage.

30. Suppose that $P(A) = 0.42$, $P(B) = 0.38$, and $P(A \text{ or } B) = 0.70$. Are A and B mutually exclusive? Explain.

31. The probability of event A occurring is $m/(2m + n)$, and the probability of event B occurring is $n/(2m + n)$. Find the probability of A or B occurring if the events are mutually exclusive.

32. Events A and B are mutually exclusive with $P(A)$ equal to 0.392 and $P(A \text{ or } B)$ equal to 0.653. Find

a. $P(B)$

b. $P(\text{not } A)$

c. $P(A \text{ and } B)$

Technology Step by Step

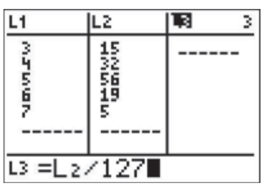
TI-84 Plus Step by Step

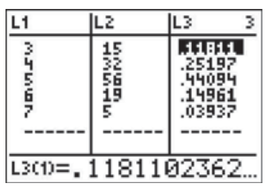
To construct a relative frequency table:

1. Enter the data values in L_1 and the frequencies in L_2.
2. Move the cursor to the top of the L_3 column so that L_3 is highlighted.
3. Type $\mathbf{L_2}$ divided by the sample size, then press **ENTER.**

Example TI4–1

Construct a relative frequency table for the knee replacement data from Example 4–14:



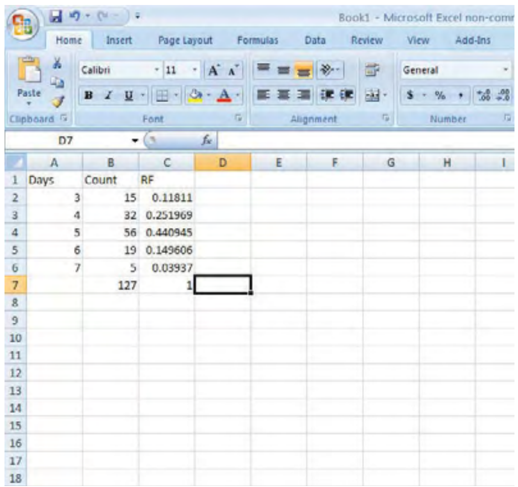


EXCEL
Step by Step

Constructing a Relative Frequency Distribution

Use the data from Example 4–14.

1. In a new worksheet, type the label **DAYS** in cell A1. Beginning in cell A2, type in the data for the variable representing the number of days knee replacement patients stayed in the hospital.
2. In cell B1, type the label for the frequency, **COUNT.** Beginning in cell B2, type in the frequencies.
3. In cell B7, compute the total frequency by selecting the sum icon Σ from the toolbar and press **Enter.**
4. In cell C1, type a label for the relative frequencies, **RF.** In cell C2, type **=(B2)/(B7)** and **Enter.** In cell C3, type =(**B3**)/(**B7**) and **Enter.** Repeat this for each of the remaining frequencies.
5. To find the total relative frequency, select the sum icon Σ from the toolbar and **Enter.** This sum should be 1.



Constructing a Contingency Table

Example XL4–1

For this example, you will need to have the MegaStat Add-In installed on Excel (refer to Chapter 1, Excel Step by Step instructions for instructions on installing MegaStat).

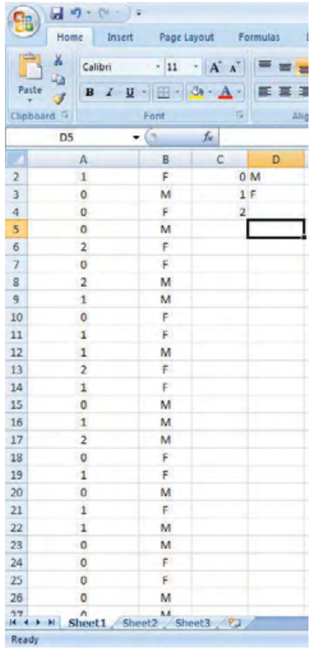
1. Open the Databank.xls file from the text website. To do this:

   Copy the files to your Desktop by choosing “Save Target As ...” or “Save Link As ...”

   Double-click the datasets folder. Then double-click the all_data-sets folder.

   Double-click the bluman_es_data-sets_excel-windows folder. In this folder double-click the Databank.xls file. The Excel program will open automatically once you open this file.
2. Highlight the column labeled SMOKING STATUS to copy these data onto a new Excel worksheet.
3. Click the Microsoft Office Button, select New Blank Workbook, then Create.
4. With cell A1 selected, click the Paste icon on the toolbar to paste the data into the new workbook.
5. Return to the Databank.xls file. Highlight the column labeled Gender. Copy and paste these data into column B of the worksheet containing the SMOKING STATUS data.

6. Type in the categories for SMOKING STATUS, **0, 1,** and **2** into cells C2–C4. In cell D2, type M for male and in cell D3, type F for female.



7. On the toolbar, select Add-Ins. Then select MegaStat. *Note:* You may need to open MegaStat from the file MegaStat.xls saved on your computer's hard drive.
8. Select **Chi-Square/Crosstab>Crosstabulation.**
9. In the Row variable Data range box, type A1:A101. In the Row variable Specification range box, type C2:C4. In the Column variable Data range box, type B1:B101. In the Column variable Specification range box, type D2:D3. Remove any checks from the Output Options. Then click [OK].

Crosstabulation

		GENDER		
		M	F	Total
SMOKING STATUS	0	21	25	46
	1	19	18	37
	2	9	8	17
	Total	49	51	100

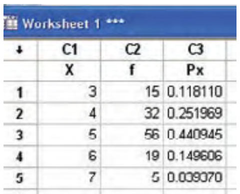
MINITAB
Step by Step

Calculate Relative Frequency Probabilities

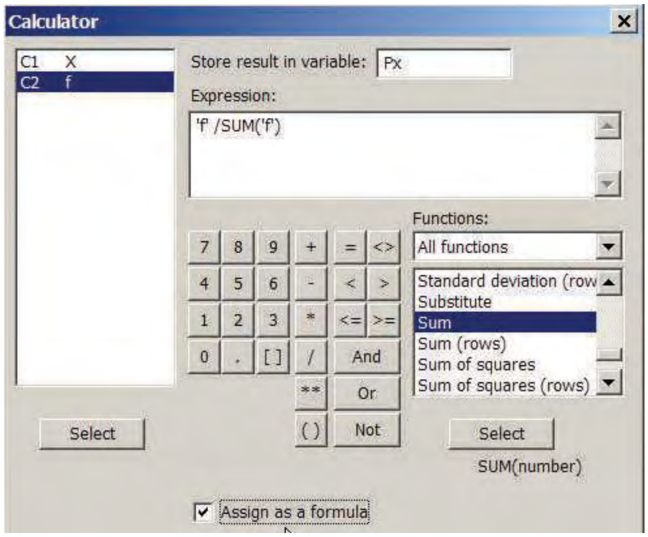
The random variable X represents the number of days patients stayed in the hospital from Example 4–14.

1. In C1 of a worksheet, type in the values of X. Name the column **X.**
2. In C2 enter the frequencies. Name the column **f.**
3. To calculate the relative frequencies and store them in a new column named Px:
   a) Select **Calc>Calculator.**
   b) Type **Px** in the box for Store result in variable:.
   c) Click in the Expression box, then double-click C2 f.
   d) Type or click the division operator.
   e) Scroll down the function list to Sum, then click [Select].
   f) Double-click C2 f to select it.
   g) Click [OK].

The dialog box and completed worksheet are shown.



↓	C1	C2	C3
	X	f	Px
1	3	15	0.118110
2	4	32	0.251969
3	5	56	0.440945
4	6	19	0.149606
5	7	5	0.039370



If the original data, rather than the table, are in a worksheet, use **Stat>Tables>Tally** to make the tables with percents (Section 2–1).

MINITAB can also make a two-way classification table.

Construct a Contingency Table

1. Select **File>Open Worksheet** to open the Databank.mtw file.
2. Select **Stat>Tables>Crosstabulation . . .**
   a) Press [TAB] and then double-click C4 SMOKING STATUS to select it for the Rows: Field.
   b) Press [TAB] and then select C11 GENDER for the Columns: Field.
   c) Click on option for Counts and then [OK].

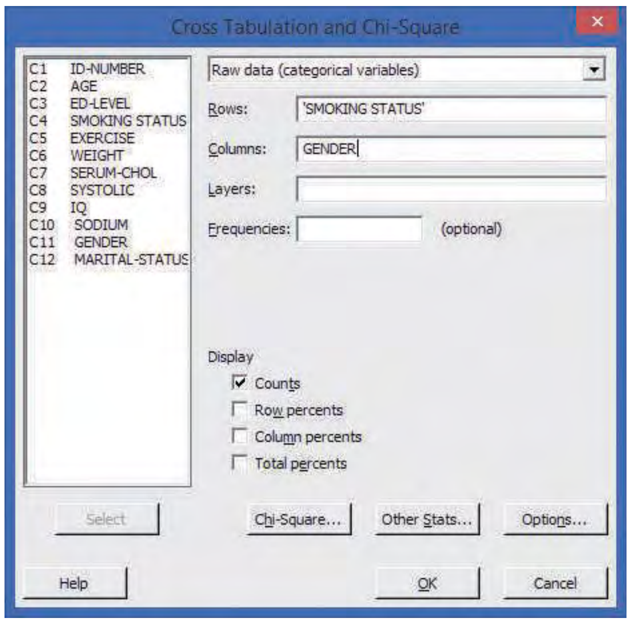
The session window and completed dialog box are shown.

Tabulated statistics: SMOKING STATUS, GENDER

Rows: SMOKING STATUS Columns: GENDER

	F	M	All
0	25	22	47
1	18	19	37
2	7	9	16
All	50	50	100

Cell Contents: Count



In this sample of 100 there are 25 females who do not smoke compared to 22 men. Sixteen individuals smoke 1 pack or more per day.

4–3 The Multiplication Rules and Conditional Probability

Section 4–2 showed that the addition rules are used to compute probabilities for mutually exclusive and non-mutually exclusive events. This section introduces the multiplication rules.

The Multiplication Rules

OBJECTIVE 3

Find the probability of compound events, using the multiplication rules.

The *multiplication rules* can be used to find the probability of two or more events that occur in sequence. For example, if you toss a coin and then roll a die, you can find the probability of getting a head on the coin *and* a 4 on the die. These two events are said to be *independent* since the outcome of the first event (tossing a coin) does not affect the probability outcome of the second event (rolling a die).

Two events A and B are **independent events** if the fact that A occurs does not affect the probability of B occurring.

Here are other examples of independent events:

Rolling a die and getting a 6, and then rolling a second die and getting a 3.

Drawing a card from a deck and getting a queen, replacing it, and drawing a second card and getting a queen.

To find the probability of two independent events that occur in sequence, you must find the probability of each event occurring separately and then multiply the answers. For example, if a coin is tossed twice, the probability of getting two heads is $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$. This result can be verified by looking at the sample space HH, HT, TH, TT. Then $P(\text{HH}) = \frac{1}{4}$.

Multiplication Rule 1

When two events are independent, the probability of both occurring is

$$P(A \text{ and } B) = P(A) \cdot P(B)$$

EXAMPLE 4–23 Tossing a Coin

A coin is flipped and a die is rolled. Find the probability of getting a head on the coin and a 4 on the die.

SOLUTION

The sample space for the coin is H, T; and for the die it is 1, 2, 3, 4, 5, 6.

$$P(\text{head and } 4) = P(\text{head}) \cdot P(4) = \frac{1}{2} \cdot \frac{1}{6} = \frac{1}{12} \approx 0.083$$

The problem in Example 4–23 can also be solved by using the sample space

H1 H2 H3 H4 H5 H6 T1 T2 T3 T4 T5 T6

The solution is $\frac{1}{12}$, since there is only one way to get the head-4 outcome.

EXAMPLE 4–24 Drawing a Card

A card is drawn from a deck and replaced; then a second card is drawn. Find the probability of getting a king and then a 7.

SOLUTION

The probability of getting a king is $\frac{4}{52}$, and the probability of getting a 7 is $\frac{4}{52}$; hence, the probability of getting a king and then a 7 is

$$P(\text{king and then a 7}) = \frac{4}{52} \cdot \frac{4}{52} = \frac{16}{2704} = \frac{1}{169} \approx 0.006.$$

EXAMPLE 4–25 Selecting a Colored Ball

An urn contains 2 red balls, 5 blue balls, and 3 white balls. A ball is selected and its color is noted. Then it is replaced. A second ball is selected and its color is noted. Find the probability of each of these events.

a. Selecting 3 blue balls

b. Selecting 1 white ball and then a red ball

c. Selecting 2 blue balls and then one white ball

SOLUTION

a. $P(\text{blue and blue and blue}) = P(\text{blue}) \cdot P(\text{blue}) \cdot P(\text{blue}) = \frac{5}{10} \cdot \frac{5}{10} \cdot \frac{5}{10} = \frac{125}{1000} = \frac{1}{8} = 0.125$

b. $P(\text{white and red}) = P(\text{white}) \cdot P(\text{red}) = \frac{3}{10} \cdot \frac{2}{10} = \frac{6}{100} = \frac{3}{50} = 0.06$

c. $P(\text{blue and blue and white}) = P(\text{blue}) \cdot P(\text{blue}) \cdot P(\text{white}) = \frac{5}{10} \cdot \frac{5}{10} \cdot \frac{3}{10} = \frac{75}{1000} = \frac{3}{40} = 0.075$

Multiplication rule 1 can be extended to three or more independent events by using the formula

$$P(A \text{ and } B \text{ and } C \text{ and } \ldots \text{ and } K) = P(A) \cdot P(B) \cdot P(C) \cdots P(K)$$

When a small sample is selected from a large population and the subjects are not replaced, the probability of the event occurring changes so slightly that for the most part, it is considered to remain the same. Examples 4–26 and 4–27 illustrate this concept.

EXAMPLE 4–26 Bank Robberies

It was found that 3 out of every 4 people who commit a bank robbery are apprehended. (*Christian Science Monitor*). If 3 bank robberies are selected at random, find the probability that all three robbers will be apprehended.

SOLUTION

Let R be the probability that a bank robber is apprehended. Then

$$P(R \text{ and } R \text{ and } R) = P(R) \cdot P(R) \cdot P(R) = \frac{3}{4} \cdot \frac{3}{4} \cdot \frac{3}{4} = \frac{27}{64} \approx 0.422$$

There is about a 42% chance that all 3 robbers will be apprehended.

EXAMPLE 4–27 Male Color Blindness

Approximately 9% of men have a type of color blindness that prevents them from distinguishing between red and green. If 3 men are selected at random, find the probability that all of them will have this type of red-green color blindness.

Source: USA TODAY.

SOLUTION

Let C denote red-green color blindness. Then

$$P(C \text{ and } C \text{ and } C) = P(C) \cdot P(C) \cdot P(C) = (0.09)(0.09)(0.09) = 0.000729$$

Hence, the rounded probability is 0.0007.

There is a 0.07% chance that all three men selected will have this type of red-green color blindness.

In Examples 4–23 through 4–27, the events were independent of one another, since the occurrence of the first event in no way affected the outcome of the second event. On the other hand, when the occurrence of the first event changes the probability of the occurrence of the second event, the two events are said to be *dependent*. For example, suppose a card is drawn from a deck and *not* replaced, and then a second card is drawn. What is the probability of selecting an ace on the first card and a king on the second card?

Before an answer to the question can be given, you must realize that the events are dependent. The probability of selecting an ace on the first draw is $\frac{4}{52}$. If that card is *not* replaced, the probability of selecting a king on the second card is $\frac{4}{51}$, since there are 4 kings and 51 cards remaining. The outcome of the first draw has affected the outcome of the second draw.

Dependent events are formally defined now.

When the outcome or occurrence of the first event affects the outcome or occurrence of the second event in such a way that the probability is changed, the events are said to be **dependent events.**

Here are some examples of dependent events:

Drawing a card from a deck, not replacing it, and then drawing a second card

Selecting a ball from an urn, not replacing it, and then selecting a second ball

Being a lifeguard and getting a suntan

Having high grades and getting a scholarship

Parking in a no-parking zone and getting a parking ticket

To find probabilities when events are dependent, use the multiplication rule with a modification in notation. For the problem just discussed, the probability of getting an ace on the first draw is $\frac{4}{52}$, and the probability of getting a king on the second draw is $\frac{4}{51}$. By the multiplication rule, the probability of both events occurring is

$$\frac{4}{52} \cdot \frac{4}{51} = \frac{16}{2652} = \frac{4}{663} \approx 0.006$$

The event of getting a king on the second draw *given* that an ace was drawn the first time is called a *conditional probability.*

The **conditional probability** of an event B in relationship to an event A is the probability that event B occurs after event A has already occurred. The notation for conditional probability is $P(B|A)$. This notation does not mean that B is divided by A; rather, it means the probability that event B occurs given that event A has already occurred. In the card example, $P(B|A)$ is the probability that the second card is a king given that the first card is an ace, and it is equal to $\frac{4}{51}$ since the first card was *not* replaced.

Multiplication Rule 2

When two events are dependent, the probability of both occurring is

$$P(A \text{ and } B) = P(A) \cdot P(B|A)$$

EXAMPLE 4–28 Unemployed Workers

For a specific year, 5.2% of U.S. workers were unemployed. During that time, 33% of those who were unemployed received unemployment benefits. If a person is selected at random, find the probability that she or he received unemployment benefits if the person is unemployed.

Source: Bureau of Labor Statistics

SOLUTION

$$P(\text{unemployed benefits and unemployed}) = P(U) \cdot P(B|U) = (0.052)(0.33) = 0.017$$

There is a 0.017 probability that a person is unemployed and receiving unemployment benefits.

EXAMPLE 4–29 Homeowner's and Automobile Insurance

World Wide Insurance Company found that 53% of the residents of a city had homeowner's insurance (H) with the company. Of these clients, 27% also had automobile insurance (A) with the company. If a resident is selected at random, find the probability that the resident has both homeowner's and automobile insurance with World Wide Insurance Company.

SOLUTION

$$P(\text{H and A}) = P(\text{H}) \cdot P(\text{A}|\text{H}) = (0.53)(0.27) = 0.1431 \approx 0.143$$

There is about a 14.3% probability that a resident has both homeowner's and automobile insurance with World Wide Insurance Company.

This multiplication rule can be extended to three or more events, as shown in Example 4–30.

EXAMPLE 4–30 Drawing Cards

Three cards are drawn from an ordinary deck and not replaced. Find the probability of these events.

a. Getting 3 jacks

b. Getting an ace, a king, and a queen in order

c. Getting a club, a spade, and a heart in order

d. Getting 3 clubs

SOLUTION

a. $P(\text{3 jacks}) = \dfrac{4}{52} \cdot \dfrac{3}{51} \cdot \dfrac{2}{50} = \dfrac{24}{132{,}600} = \dfrac{1}{5525} \approx 0.0002$

b. $P(\text{ace and king and queen}) = \dfrac{4}{52} \cdot \dfrac{4}{51} \cdot \dfrac{4}{50} = \dfrac{64}{132{,}600} = \dfrac{8}{16{,}575} \approx 0.0005$

c. $P(\text{club and spade and heart}) = \dfrac{13}{52} \cdot \dfrac{13}{51} \cdot \dfrac{13}{50} = \dfrac{2197}{132{,}600} = \dfrac{169}{10{,}200} \approx 0.017$

d. $P(\text{3 clubs}) = \dfrac{13}{52} \cdot \dfrac{12}{51} \cdot \dfrac{11}{50} = \dfrac{1716}{132{,}600} = \dfrac{11}{850} \approx 0.013$

Tree diagrams can be used as an aid to finding the solution to probability problems when the events are sequential. Example 4–31 illustrates the use of tree diagrams.

EXAMPLE 4–31 Selecting Colored Balls

Box 1 contains 2 red balls and 1 blue ball. Box 2 contains 3 blue balls and 1 red ball. A coin is tossed. If it falls heads up, box 1 is selected and a ball is drawn. If it falls tails up, box 2 is selected and a ball is drawn. Find the probability of selecting a red ball.

SOLUTION

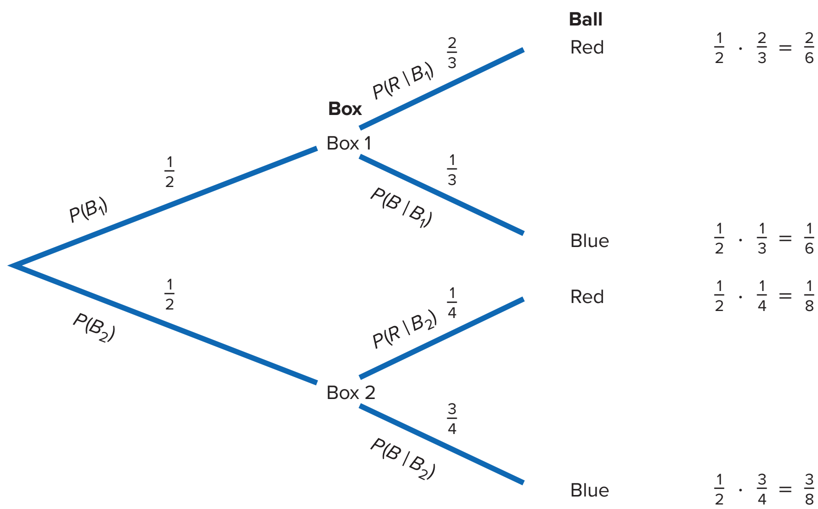
The first two branches designate the selection of either box 1 or box 2. Then from box 1, either a red ball or a blue ball can be selected. Likewise, a red ball or blue ball can be selected from box 2. Hence, a tree diagram of the example is shown in Figure 4–8.

Next determine the probabilities for each branch. Since a coin is being tossed for the box selection, each branch has a probability of $\frac{1}{2}$, that is, heads for box 1 or tails for box 2. The probabilities for the second branches are found by using the basic probability rule. For example, if box 1 is selected and there are 2 red balls and 1 blue ball, the probability of selecting a red ball is $\frac{2}{3}$ and the probability of selecting a blue ball is $\frac{1}{3}$. If box 2 is selected and it contains 3 blue balls and 1 red ball, then the probability of selecting a red ball is $\frac{1}{4}$ and the probability of selecting a blue ball is $\frac{3}{4}$.

Next multiply the probability for each outcome, using the rule $P(A \text{ and } B) = P(A) \cdot P(B|A)$. For example, the probability of selecting box 1 and selecting a red ball is $\frac{1}{2} \cdot \frac{2}{3} = \frac{2}{6}$. The probability of selecting box 1 and a blue ball is $\frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$. The probability of selecting box 2 and selecting a red ball is $\frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}$. The probability of selecting box 2 and a blue ball is $\frac{1}{2} \cdot \frac{3}{4} = \frac{3}{8}$. (Note that the sum of these probabilities is 1.)

Finally, a red ball can be selected from either box 1 or box 2 so $P(\text{red}) = \frac{2}{6} + \frac{1}{8} = \frac{8}{24} + \frac{3}{24} = \frac{11}{24}$.

FIGURE 4–8 Tree Diagram for Example 4–31



Tree diagrams can be used when the events are independent or dependent, and they can also be used for sequences of three or more events.

OBJECTIVE 4

Find the conditional probability of an event.

Conditional Probability

The conditional probability of an event B in relationship to an event A was defined as the probability that event B occurs after event A has already occurred.

The conditional probability of an event can be found by dividing both sides of the equation for multiplication rule 2 by $P(A)$, as shown:

$$P(A \text{ and } B) = P(A) \cdot P(B|A)$$

$$\frac{P(A \text{ and } B)}{P(A)} = \frac{\cancel{P(A)} \cdot P(B|A)}{\cancel{P(A)}}$$

$$\frac{P(A \text{ and } B)}{P(A)} = P(B|A)$$

Formula for Conditional Probability

The probability that the second event B occurs given that the first event A has occurred can be found by dividing the probability that both events occurred by the probability that the first event has occurred. The formula is

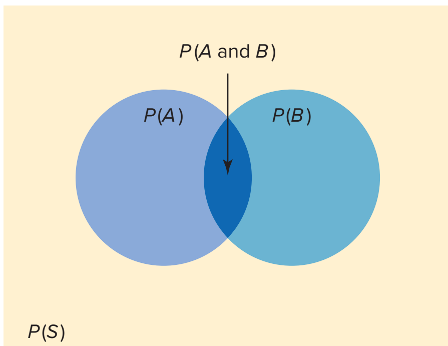
$$P(B|A) = \frac{P(A \text{ and } B)}{P(A)}$$

The Venn diagram for conditional probability is shown in Figure 4–9. In this case,

$$P(B|A) = \frac{P(A \text{ and } B)}{P(A)}$$

which is represented by the area in the intersection or overlapping part of the circles A and B, divided by the area of circle A. The reasoning here is that if you assume A has occurred, then A becomes the sample space for the next calculation and is the denominator of the probability fraction $P(A \text{ and } B)/P(A)$. The numerator $P(A \text{ and } B)$ represents the probability of the part of B that is contained in A. Hence, $P(A \text{ and } B)$ becomes the numerator of the probability fraction $P(A \text{ and } B)/P(A)$. Imposing a condition reduces the sample space.

FIGURE 4–9
Venn Diagram for Conditional Probability



$$P(B|A) = \frac{P(A \text{ and } B)}{P(A)}$$

Examples 4–32, 4–33, and 4–34 illustrate the use of this rule.

EXAMPLE 4–32 Selecting Colored Chips

A box contains black chips and white chips. A person selects two chips without replacement. If the probability of selecting a black chip *and* a white chip is $\frac{15}{56}$ and the probability of selecting a black chip on the first draw is $\frac{3}{8}$, find the probability of selecting the white chip on the second draw, *given* that the first chip selected was a black chip.

SOLUTION

Let

$$B = \text{selecting a black chip} \qquad W = \text{selecting a white chip}$$

Then

$$P(W|B) = \frac{P(B \text{ and } W)}{P(B)} = \frac{15/56}{3/8}$$

$$= \frac{15}{56} \div \frac{3}{8} = \frac{15}{56} \cdot \frac{8}{3} = \frac{\overset{5}{\cancel{15}}}{\underset{7}{\cancel{56}}} \cdot \frac{\overset{1}{\cancel{8}}}{\underset{1}{\cancel{3}}} = \frac{5}{7} \approx 0.714$$

Hence, the probability of selecting a white chip on the second draw given that the first chip selected was black is $\frac{5}{7} \approx 0.714$.

EXAMPLE 4–33 Parking Tickets

The probability that Sam parks in a no-parking zone *and* gets a parking ticket is 0.06, and the probability that Sam cannot find a legal parking space and has to park in the no-parking zone is 0.20. On Tuesday, Sam arrives at school and has to park in a no-parking zone. Find the probability that he will get a parking ticket.

SOLUTION

Let

$$N = \text{parking in a no-parking zone} \qquad T = \text{getting a ticket}$$

Then

$$P(T|N) = \frac{P(N \text{ and } T)}{P(N)} = \frac{0.06}{0.20} = 0.30$$

Hence, Sam has a 0.30 probability or 30% chance of getting a parking ticket, given that he parked in a no-parking zone.

The conditional probability of events occurring can also be computed when the data are given in table form, as shown in Example 4–34.

EXAMPLE 4–34 Survey on Women in the Military

A recent survey asked 100 people if they thought women in the armed forces should be permitted to participate in combat. The results of the survey are shown.

Gender	Yes	No	Total
Male	32	18	50
Female	8	42	50
Total	40	60	100

Find these probabilities.

a. The respondent answered yes, given that the respondent was a female.

b. The respondent was a male, given that the respondent answered no.

SOLUTION

Let

M = respondent was a male $\quad$ Y = respondent answered yes

F = respondent was a female $\quad$ N = respondent answered no

a. The problem is to find $P(Y|F)$. The rule states

$$P(Y|F) = \frac{P(F \text{ and } Y)}{P(F)}$$

The probability $P(F \text{ and } Y)$ is the number of females who responded yes, divided by the total number of respondents:

$$P(F \text{ and } Y) = \frac{8}{100}$$

The probability $P(F)$ is the probability of selecting a female:

$$P(F) = \frac{50}{100}$$

Then

$$P(Y|F) = \frac{P(F \text{ and } Y)}{P(F)} = \frac{8/100}{50/100}$$

$$= \frac{8}{100} \div \frac{50}{100} = \frac{\overset{4}{\cancel{8}}}{\underset{1}{\cancel{100}}} \cdot \frac{\overset{1}{\cancel{100}}}{\underset{25}{\cancel{50}}} = \frac{4}{25} = 0.16$$

b. The problem is to find $P(M|N)$.

$$P(M|N) = \frac{P(N \text{ and } M)}{P(N)} = \frac{18/100}{60/100}$$

$$= \frac{18}{100} \div \frac{60}{100} = \frac{\overset{3}{\cancel{18}}}{\underset{1}{\cancel{100}}} \cdot \frac{\overset{1}{\cancel{100}}}{\underset{10}{\cancel{60}}} = \frac{3}{10} = 0.3$$

Probabilities for "At Least"

The multiplication rules can be used with the complementary event rule (Section 4–1) to simplify solving probability problems involving "at least." Examples 4–35, 4–36, and 4–37 illustrate how this is done.

EXAMPLE 4–35 Drawing Cards

A person selects 3 cards from an ordinary deck and replaces each card after it is drawn. Find the probability that the person will get at least one heart.

SOLUTION

It is much easier to find the probability that the person will not select a heart in three draws and subtract this value from 1. To do the problem directly, you would have to find the probability of selecting 1 heart, 2 hearts, and 3 hearts and then add the results.

Let

$$E = \text{at least 1 heart is drawn} \qquad \text{and} \qquad \overline{E} = \text{no hearts are drawn}$$

$$P(\overline{E}) = \frac{39}{52} \cdot \frac{39}{52} \cdot \frac{39}{52} = \frac{3}{4} \cdot \frac{3}{4} \cdot \frac{3}{4} = \frac{27}{64}$$

$$P(E) = 1 - P(\overline{E})$$

$$= 1 - \frac{27}{64} = \frac{37}{64} \approx 0.578 = 57.8\%$$

Hence, a person will select at least one heart about 57.8% of the time.

EXAMPLE 4–36 Rolling a Die

A single die is rolled 4 times. Find the probability of getting at least one 6.

SOLUTION

It is easier to find the probability of the complement of the event, which is no 6s. Then subtract this probability from 1 in order to find the probability of getting at least one 6.

$$P(\text{at least one 6}) = 1 - P(\text{no 6s})$$

$$= 1 - \left(\frac{5}{6}\right)^4$$

$$= 1 - \frac{625}{1296}$$

$$= \frac{671}{1296} \approx 0.518$$

There is about a 51.8% chance of getting at least one 6 when a die is rolled four times.

EXAMPLE 4–37 Ties

The Neckware Association of America reported that 3% of ties sold in the United States are bow ties. If 4 customers who purchased a tie are randomly selected, find the probability that at least 1 purchased a bow tie.

SOLUTION

Let E = at least 1 bow tie is purchased and $\overline{E}$ = no bow ties are purchased. Then

$$P(E) = 0.03 \qquad \text{and} \qquad P(\overline{E}) = 1 - 0.03 = 0.97$$

P(no bow ties are purchased) = (0.97)(0.97)(0.97)(0.97) ≈ 0.885; hence,
P(at least one bow tie is purchased) = 1 − 0.885 = 0.115.

There is an 11.5% chance of a person purchasing at least one bow tie.

Applying the Concepts 4–3

Guilty or Innocent?

In July 1964, an elderly woman was mugged in Costa Mesa, California. In the vicinity of the crime a tall, bearded man sat waiting in a yellow car. Shortly after the crime was committed, a young, tall woman, wearing her blond hair in a ponytail, was seen running from the scene of the crime and getting into the car, which sped off. The police broadcast a description of the suspected muggers. Soon afterward, a couple fitting the description was arrested and convicted of the crime. Although the evidence in the case was largely circumstantial, the two people arrested were nonetheless

convicted of the crime. The prosecutor based his entire case on basic probability theory, showing the unlikeness of another couple being in that area while having all the same characteristics that the elderly woman described. The following probabilities were used.

Characteristic	Assumed probability
Drives yellow car	1 out of 12
Man over 6 feet tall	1 out of 10
Man wearing tennis shoes	1 out of 4
Man with beard	1 out of 11
Woman with blond hair	1 out of 3
Woman with hair in a ponytail	1 out of 13
Woman over 6 feet tall	1 out of 100

1. Compute the probability of another couple being in that area with the same characteristics.
2. Would you use the addition or multiplication rule? Why?
3. Are the characteristics independent or dependent?
4. How are the computations affected by the assumption of independence or dependence?
5. Should any court case be based solely on probabilities?
6. Would you convict the couple who was arrested even if there were no eyewitnesses?
7. Comment on why in today's justice system no person can be convicted solely on the results of probabilities.
8. In actuality, aren't most court cases based on uncalculated probabilities?

See page 254 for the answers.

Exercises 4–3

1. State which events are independent and which are dependent.

a. Tossing a coin and drawing a card from a deck

b. Drawing a ball from an urn, not replacing it, and then drawing a second ball

c. Getting a raise in salary and purchasing a new car

d. Driving on ice and having an accident

2. State which events are independent and which are dependent.

a. Having a large shoe size and having a high IQ

b. A father being left-handed and a daughter being left-handed

c. Smoking excessively and having lung cancer

d. Eating an excessive amount of ice cream and smoking an excessive amount of cigarettes

3. Video and Computer Games Sixty-nine percent of U.S. heads of household play video or computer games. Choose 4 heads of household at random. Find the probability that

a. None play video or computer games.

b. All four do.

Source: www.theesa.com

4. Seat Belt Use The Gallup Poll reported that 52% of Americans used a seat belt the last time they got into a car. If 4 people are selected at random, find the probability that they all used a seat belt the last time they got into a car.

Source: 100% American.

5. Automobile Sales An automobile salesperson finds the probability of making a sale is 0.21. If she talks to 4 customers, find the probability that she will make 4 sales. Is the event likely or unlikely to occur? Explain your answer.

6. Prison Populations If 25% of U.S. federal prison inmates are not U.S. citizens, find the probability that 2 randomly selected federal prison inmates will not be U.S. citizens.

Source: Harper's Index.

7. Government Employees In 2013 about 66% of full-time law enforcement workers were sworn officers, and of those, 88.4% were male. Females however make up 60.7% of civilian employees. Choose one law enforcement worker at random and find the following.

a. The probability that she is a female sworn officer

b. The probability that he is a male civilian employee

c. The probability that he or she is male or a civilian employee

Source: World Almanac.

8. Working Women and Computer Use It is reported that 72% of working women use computers at work. Choose 5 working women at random. Find

a. The probability that at least 1 doesn't use a computer at work

b. The probability that all 5 use a computer in their jobs

Source: www.infoplease.com

9. Female Prison Inmates Seventy-five percent of female prison inmates are mothers. If 3 female prison inmates are selected at random, what is the probability that none are mothers?

Source: Chicago Legal Aid to Incarcerated Mothers.

10. Selecting Marbles A bag contains 9 red marbles, 8 white marbles, and 6 blue marbles. Randomly choose two marbles, one at a time, and without replacement. Find the following.

a. The probability that the first marble is red and the second is white

b. The probability that both are the same color

c. The probability that the second marble is blue

11. Smart TVs Smart TVs have seen success in the United States market. During the 2nd quarter of 2015 45% of TVs sold in the United States were Smart TVs. That's an increase of 11% from 2014. Choose three households and find the probability that

a. None of the 3 households had a Smart TV

b. All 3 households had a Smart TV

c. At least 1 of the 3 households had a Smart TV

Source: The NPD Group Connected Intelligence Home Entertainment Report

12. Flashlight Batteries A flashlight has 6 batteries, 2 of which are defective. If 2 are selected at random without replacement, find the probability that both are defective.

13. Drawing a Card Four cards are drawn from a desk *without* replacement. Find these probabilities.

a. All cards are jacks.

b. All cards are black cards.

c. All cards are hearts.

14. Scientific Study In a scientific study there are 8 guinea pigs, 5 of which are pregnant. If 3 are selected at random without replacement, find the probability that all are pregnant.

15. Drawing Cards If two cards are selected from a standard deck of 52 cards and are not replaced after each draw, find these probabilities.

a. Both are 9s.

b. Both cards are the same suit.

c. Both cards are spades.

16. Winning a Door Prize At a gathering consisting of 10 men and 20 women, two door prizes are awarded. Find the probability that both prizes are won by men. The winning ticket is not replaced. Would you consider this event likely or unlikely to occur?

17. Defective Batteries In a box of 12 batteries, 2 are dead. If 2 batteries are selected at random for a flashlight, find the probability that both are dead. Would you consider this event likely or unlikely?

18. Sales A manufacturer makes two models of an item: model I, which accounts for 80% of unit sales, and model II, which accounts for 20% of unit sales. Because of defects, the manufacturer has to replace (or exchange) 10% of its model I and 18% of its model II. If a model is selected at random, find the probability that it will be defective.

19. Student Financial Aid In a recent year 8,073,000 male students and 10,980,000 female students were enrolled as undergraduates. Receiving aid were 60.6% of the male students and 65.2% of the female students. Of those receiving aid, 44.8% of the males got federal aid and 50.4% of the females got federal aid. Choose 1 student at random. (*Hint:* Make a tree diagram.) Find the probability that the student is

a. A male student without aid

b. A male student, given that the student has aid

c. A female student or a student who receives federal aid

Source: www.nces.gov

20. Selecting Colored Balls Urn 1 contains 5 red balls and 3 black balls. Urn 2 contains 3 red balls and 1 black ball. Urn 3 contains 4 red balls and 2 black balls. If an urn is selected at random and a ball is drawn, find the probability it will be red.

21. Automobile Insurance An insurance company classifies drivers as low-risk, medium-risk, and high-risk. Of those insured, 60% are low-risk, 30% are medium-risk, and 10% are high-risk. After a study, the company finds that during a 1-year period, 1% of the low-risk drivers had an accident, 5% of the medium-risk drivers had an accident, and 9% of the high-risk drivers had an accident. If a driver is selected at random, find the probability that the driver will have had an accident during the year.

22. Defective Items A production process produces an item. On average, 15% of all items produced are defective. Each item is inspected before being shipped, and the inspector misclassifies an item 10% of the time. What proportion of the items will be "classified as good"? What is the probability that an item is defective given that it was classified as good?

23. Prison Populations For a recent year, 0.99 of the incarcerated population is adults and 0.07 of the incarcerated are adult females. If an incarcerated person is selected at random, find the probability that the person is a female given that the person is an adult.

Source: Bureau of Justice.

24. Rolling Dice Roll two standard dice and add the numbers. What is the probability of getting a number larger than 9 for the first time on the third roll?

25. Heart Disease Twenty-five percent of all deaths (all ages) are caused by diseases of the heart. Ischemic heart disease accounts for 16.4% of all deaths and heart failure for 2.3%. Choose one death at random. What is the probability that it is from ischemic heart disease given that it was from heart disease? Choose two deaths at random; what is the probability that at least one is from heart disease?

Source: Time Almanac.

26. Country Club Activities At the Avonlea Country Club, 73% of the members play bridge and swim, and 82% play bridge. If a member is selected at random, find the probability that the member swims, given that the member plays bridge.

27. College Courses At a large university, the probability that a student takes calculus and is on the dean's list is 0.042. The probability that a student is on the dean's list is 0.21. Find the probability that the student is taking calculus, given that he or she is on the dean's list.

28. Congressional Terms Below is given the summary from the 112th Congress of Senators whose terms end in 2013, 2015, or 2017.

	2013	2015	2017
Democrat	21	20	1
Republican	8	15	13

Choose one of these Senators at random and find

a. *P*(Democrat and term expires in 2015)

b. *P*(Republican or term expires in 2013)

c. *P*(Republican given term expires in 2017)

Are the events "Republican" and "term expires in 2015" independent? Explain.

Source: Time Almanac 2012.

29. Pizza and Salads In a pizza restaurant, 95% of the customers order pizza. If 65% of the customers order pizza and a salad, find the probability that a customer who orders pizza will also order a salad.

30. Gift Baskets The Gift Basket Store had the following premade gift baskets containing the following combinations in stock.

	Cookies	Mugs	Candy
Coffee	20	13	10
Tea	12	10	12

Choose 1 basket at random. Find the probability that it contains

a. Coffee or candy

b. Tea given that it contains mugs

c. Tea and cookies

Source: www.infoplease.com

31. Blood Types and Rh Factors In addition to being grouped into four types, human blood is grouped by its Rhesus (Rh) factor. Consider the figures below which show the distributions of these groups for Americans.

	O	A	B	AB
Rh+	37%	34%	10%	4%
Rh−	6	6	2	1

Choose one American at random. Find the probability that the person

a. Is a universal donor, i.e., has O-negative blood

b. Has type O blood given that the person is Rh+

c. Has A+ or AB− blood

d. Has Rh− given that the person has type B

Source: www.infoplease.com

32. Doctor Specialties Below are listed the numbers of doctors in various specialties by gender.

	Pathology	Pediatrics	Psychiatry
Male	12,575	33,020	27,803
Female	5,604	33,351	12,292

Choose one doctor at random.

a. Find P (male|pediatrician).

b. Find P (pathologist|female).

c. Are the characteristics "female" and "pathologist" independent? Explain.

Source: World Almanac.

33. Lightning Strikes It has been said that the probability of being struck by lightning is about 1 in 750,000, but under what circumstances? Below are listed the numbers of deaths from lightning since 1996.

	Golf/ball field	Boating/in water	Outside/camping	Construction	Under a tree	Phone	Other
1996–2000	16	23	117	9	40	0	30
2001–2005	17	16	112	3	35	0	23
2006–2010	15	17	91	0	42	1	16

Choose one fatality at random and find each probability.

a. Given that the death was after 2000, what is the probability that it occurred under a tree?

b. Find the probability that the death was from camping or being outside and was before 2001.

c. Find the probability that the death was from camping or being outside given that it was before 2001.

Source: Noaa.gov/hazstats

34. Foreign Adoptions The following foreign adoptions (in the United States) occurred during these particular years.

	2014	2013
China	2040	2306
Ethiopia	716	993
Ukraine	521	438

Choose one adoption at random from this chart.

a. What is the probability that it was from Ethiopia given that it was from 2013?

b. What is the probability that it was from the Ukraine and in 2014?

c. What is the probability that it did not occur in 2014 and was not from Ethiopia?

d. Choose two adoptions at random; what is the probability that they were both from China?

Source: World Almanac.

35. Leisure Time Exercise Only 27% of U.S. adults get enough leisure time exercise to achieve cardiovascular fitness. Choose 3 adults at random. Find the probability that

a. All 3 get enough daily exercise

b. At least 1 of the 3 gets enough exercise

Source: www.infoplease.com

36. Customer Purchases In a department store there are 120 customers, 90 of whom will buy at least 1 item. If 5 customers are selected at random, one by one, find the probability that all will buy at least 1 item.

37. Marital Status of Women According to the *Statistical Abstract of the United States,* 70.3% of females ages 20 to 24 have never been married. Choose 5 young women in this age category at random. Find the probability that

a. None has ever been married

b. At least 1 has been married

Source: New York Times Almanac.

38. Fatal Accidents The American Automobile Association (AAA) reports that of the fatal car and truck accidents, 54% are caused by car driver error. If 3 accidents are chosen at random, find the probability that

a. All are caused by car driver error

b. None is caused by car driver error

c. At least 1 is caused by car driver error

Source: AAA quoted on CNN.

39. On-Time Airplane Arrivals According to FlightStats report released April 2013, the Salt Lake City airport led major U.S. airports in on-time arrivals with an 85.5% on-time rate. Choose 5 arrivals at random and find the probability that at least 1 was not on time.

Source: FlightStats

40. On-Time Flights A flight from Pittsburgh to Charlotte has a 90% on-time record. From Charlotte to Jacksonville, North Carolina, the flight is on time 80% of the time. The return flight from Jacksonville to Charlotte is on time 50% of the time and from Charlotte to Pittsburgh, 90% of the time. Consider a round trip from Pittsburgh to Jacksonville on these flights. Assume the flights are independent.

a. What is the probability that all 4 flights are on time?

b. What is the probability that at least 1 flight is not on time?

c. What is the probability that at least 1 flight is on time?

d. Which events are complementary?

41. Reading to Children Fifty-eight percent of American children (ages 3 to 5) are read to every day by someone at home. Suppose 5 children are randomly selected. What is the probability that at least 1 is read to every day by someone at home?

Source: Federal Interagency Forum on Child and Family Statistics.

42. Doctoral Assistantships Of Ph.D. students, 60% have paid assistantships. If 3 students are selected at random, find the probabilities that

a. All have assistantships

b. None has an assistantship

c. At least 1 has an assistantship

Source: U.S. Department of Education, *Chronicle of Higher Education.*

43. Drawing Cards If 5 cards are drawn at random from a deck of 52 cards and are not replaced, find the probability of getting at least one diamond.

44. Autism In recent years it was thought that approximately 1 in 110 children exhibited some form of autism. The most recent CDC study concluded that the proportion may be as high as 1 in 88. If indeed these new figures are correct, choose 3 children at random and find these probabilities.

a. What is the probability that none have autism?

b. What is the probability that at least 1 has autism?

c. Choose 10 children at random. What is the probability that at least 1 has autism?

Source: cdc.gov

45. Video Game. Video games are rated according to the content. The average age of a gamer is 35 years old. In 2015, 15.5% of the video games were rated Mature. Choose 5 purchased games at random. Find the probability that

a. None of the five were rated mature.

b. At least 1 of the 5 was rated mature.

Source: Nobullying.com

46. Medication Effectiveness A medication is 75% effective against a bacterial infection. Find the probability that if 12 people take the medication, at least 1 person's infection will not improve.

47. Selecting Digits If 3 digits are selected at random with replacement, find the probability of getting at least one odd number. Would you consider this event likely or unlikely? Why?

48. Selecting a Letter of the Alphabet If 3 letters of the alphabet are selected at random, find the probability of getting at least 1 letter x. Letters can be used more than once. Would you consider this event likely to happen? Explain your answer.

49. Rolling a Die A die is rolled twice. Find the probability of getting at least one 6.

50. U.S. Organ Transplants As of June 2015, 81.4% of patients were waiting on a kidney, 11.7% were waiting on a liver, and 3.1% were waiting on a heart. Choose 6 patients on the transplant waiting list at random in 2015. Find the probability that

a. All were waiting for a kidney.

b. None were waiting for a kidney.

c. At least 1 was waiting for a kidney

51. Lucky People Twelve percent of people in Western countries consider themselves lucky. If 3 people are selected at random, what is the probability that at least one will consider himself lucky?

Source: San Diego-Tribune.

52. Selecting a Flower In a large vase, there are 8 roses, 5 daisies, 12 lilies, and 9 orchids. If 4 flowers are selected at random, and not replaced, find the probability that at least 1 of the flowers is a rose. Would you consider this event likely to occur? Explain your answer.

Extending the Concepts

53. Let A and B be two mutually exclusive events. Are A and B independent events? Explain your answer.

54. Types of Vehicles The Bargain Auto Mall has the following cars in stock.

	SUV	Compact	Mid-sized
Foreign	20	50	20
Domestic	65	100	45

Are the events "compact" and "domestic" independent? Explain.

55. College Enrollment An admissions director knows that the probability a student will enroll after a campus visit is 0.55, or $P(E) = 0.55$. While students are on campus visits, interviews with professors are arranged. The admissions director computes these conditional probabilities for students enrolling after visiting three professors, DW, LP, and MH.

$P(E|\text{DW}) = 0.95 \qquad P(E|\text{LP}) = 0.55 \qquad P(E|\text{MH}) = 0.15$

Is there something wrong with the numbers? Explain.

56. Commercials Event A is the event that a person remembers a certain product commercial. Event B is the event that a person buys the product. If $P(B) = 0.35$, comment on each of these conditional probabilities if you were vice president for sales.

a. $P(B|A) = 0.20$

b. $P(B|A) = 0.35$

c. $P(B|A) = 0.55$

57. Given a sample space with events A and B such that $P(A) = 0.342$, $P(B) = 0.279$, and $P(A \text{ or } B) = 0.601$. Are A and B mutually exclusive? Are A and B independent? Find $P(A|B)$, $P(\text{not } B)$, and $P(A \text{ and } B)$.

58. Child's Board Game In a child's board game of the tortoise and the hare, the hare moves by roll of a standard die and the tortoise by a six-sided die with the numbers 1, 1, 1, 2, 2, and 3. Roll each die once. What is the probability that the tortoise moves ahead of the hare?

59. Bags Containing Marbles Two bags contain marbles. Bag 1 contains 1 black marble and 9 white marbles. Bag 2 contains 1 black marble and x white marbles. If you choose a bag at random, then choose a marble at random, the probability of getting a black marble is $\frac{2}{15}$. How many white marbles are in bag 2?

4–4 Counting Rules

Many times a person must know the number of all possible outcomes for a sequence of events. To determine this number, three rules can be used: the *fundamental counting rule,* the *permutation rule,* and the *combination rule.* These rules are explained here, and they will be used in Section 4–5 to find probabilities of events.

The first rule is called the **fundamental counting rule.**

The Fundamental Counting Rule

OBJECTIVE 5

Find the total number of outcomes in a sequence of events, using the fundamental counting rule.

Fundamental Counting Rule

In a sequence of n events in which the first one has k_1 possibilities and the second event has k_2 and the third has k_3, and so forth, the total number of possibilities of the sequence will be

$$k_1 \cdot k_2 \cdot k_3 \cdots k_n$$

Note: In this case *and* means to multiply.

Examples 4–38 through 4–41 illustrate the fundamental counting rule.

EXAMPLE 4–38 Tossing a Coin and Rolling a Die

A coin is tossed and a die is rolled. Find the number of outcomes for the sequence of events.

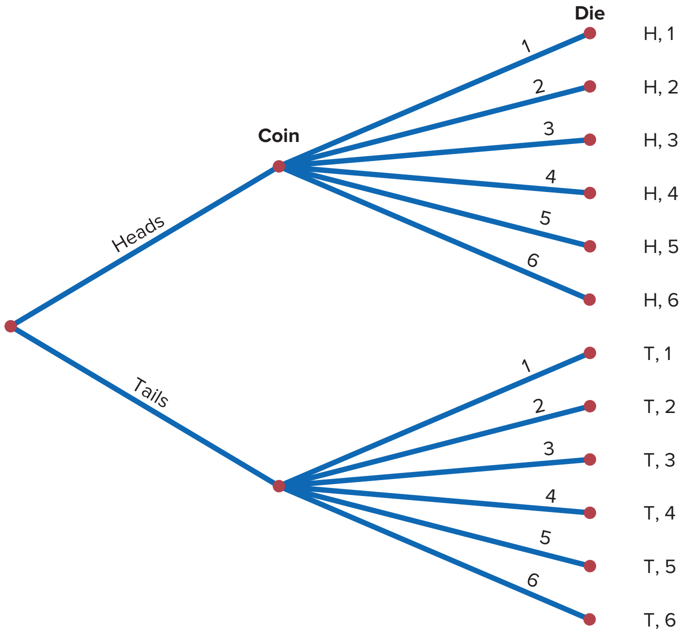
Interesting Fact

Possible games of chess: 25×10^{115}.

SOLUTION

Since the coin can land either heads up or tails up and since the die can land with any one of six numbers showing face up, there are $2 \cdot 6 = 12$ possibilities. A tree diagram can also be drawn for the sequence of events. See Figure 4–10.

FIGURE 4–10
Complete Tree Diagram for Example 4–38



EXAMPLE 4–39 Types of Paint

A paint manufacturer wishes to manufacture several different paints. The categories include

Color	Red, blue, white, black, green, brown, yellow
Type	Latex, oil
Texture	Flat, semigloss, high gloss
Use	Outdoor, indoor

How many different kinds of paint can be made if you can select one color, one type, one texture, and one use?

SOLUTION

You can choose one color and one type and one texture and one use. Since there are 7 color choices, 2 type choices, 3 texture choices, and 2 use choices, the total number of possible different paints is as follows:

Color		Type		Texture		Use	
$\boxed{7}$	·	$\boxed{2}$	·	$\boxed{3}$	·	$\boxed{2}$	= 84

EXAMPLE 4–40 Distribution of Blood Types

There are four blood types, A, B, AB, and O. Blood can also be Rh+ and Rh−. Finally, a blood donor can be classified as either male or female. How many different ways can a donor have his or her blood labeled?

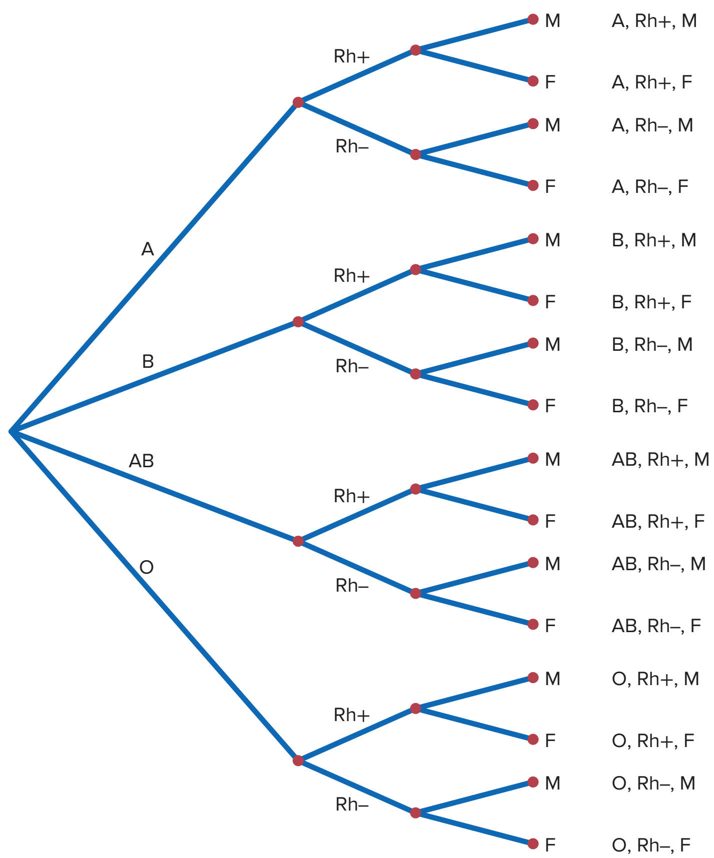
SOLUTION

Since there are 4 possibilities for blood type, 2 possibilities for Rh factor, and 2 possibilities for the gender of the donor, there are 4 · 2 · 2, or 16, different classification categories, as shown.

Blood type		Rh		Gender	
$\boxed{4}$	·	$\boxed{2}$	·	$\boxed{2}$	= 16

A tree diagram for the events is shown in Figure 4–11.

FIGURE 4–11
Complete Tree Diagram for Example 4–40



When determining the number of different possibilities of a sequence of events, you must know whether repetitions are permissible.

EXAMPLE 4–41 Railroad Memorial License Plates

The first year the state of Pennsylvania issued railroad memorial license plates, the plates had a picture of a steam engine followed by four digits. Assuming that repetitions are allowed, how many railroad memorial plates could be issued?

SOLUTION

Since there are four spaces to fill for each space, the total number of plates that can be issued is $10 \cdot 10 \cdot 10 \cdot 10 = 10{,}000$. *Note:* Actually there was such a demand for the plates, Pennsylvania had to use letters also.

Now if repetitions are not permitted, the first digit in the plate in Example 4–41 could be selected in 10 ways, the second digit in 9 ways, etc. So the total number of plates that could be issued is $10 \cdot 9 \cdot 8 \cdot 7 = 5040$.

The same situation occurs when one is drawing balls from an urn or cards from a deck. If the ball or card is replaced before the next one is selected, then repetitions are permitted, since the same one can be selected again. But if the selected ball or card is not replaced, then repetitions are not permitted, since the same ball or card cannot be selected the second time.

These examples illustrate the fundamental counting rule. In summary: *If repetitions are permitted, then the numbers stay the same going from left to right. If repetitions are not permitted, then the numbers decrease by* 1 *for each place left to right.*

Two other rules that can be used to determine the total number of possibilities of a sequence of events are the permutation rule and the combination rule.

Factorial Notation

Historical Note

In 1808 Christian Kramp first used factorial notation.

These rules use *factorial notation.* The factorial notation uses the exclamation point.

$$5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$$
$$9! = 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$$

To use the formulas in the permutation and combination rules, a special definition of 0! is needed: $0! = 1$.

Factorial Formulas

For any counting number n

$$n! = n(n-1)(n-2)\cdots 1$$
$$0! = 1$$

Permutations

A **permutation** is an arrangement of n objects in a specific order.

Examples 4–42 and 4–43 illustrate permutations.

EXAMPLE 4–42 Lock Codes

A person needs to make up a four-digit combination to open the lock on the garage door. The person decides to use four different digits from the first four digits, 0 through 3. How many different combinations can be made?

SOLUTION

The person has four choices for the first digit, three ways for the second digit, etc. Hence,

$$4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$$

There are 24 different combinations that can be used.

In Example 4–42 all objects were used up. But what happens when not all objects are used up? The answer to this question is given in Example 4–43.

EXAMPLE 4–43 Business Location

A business owner wishes to rank the top 3 locations selected from 5 locations for a business. How many different ways can she rank them?

SOLUTION

Using the fundamental counting rule, she can select any one of the 5 for first choice, then any one of the remaining 4 locations for her second choice, and finally, any one of the remaining locations for her third choice, as shown.

First choice		Second choice		Third choice	
$\boxed{5}$	$\cdot$	$\boxed{4}$	$\cdot$	$\boxed{3}$	$= 60$

The solutions in Examples 4–42 and 4–43 are permutations.

OBJECTIVE 6

Find the number of ways that r objects can be selected from n objects, using the permutation rule.

Permutation Rule 1

The arrangement of n objects in a specific order using r objects at a time is called a *permutation of n objects taking r objects at a time.* It is written as ${}_nP_r$, and the formula is

$${}_nP_r = \frac{n!}{(n-r)!}$$

The notation ${}_nP_r$ is used for permutations.

$${}_6P_4 \text{ means } \quad \frac{6!}{(6-4)!} \quad \text{or} \quad \frac{6!}{2!} = \frac{6 \cdot 5 \cdot 4 \cdot 3 \cdot \cancel{2} \cdot \cancel{1}}{\cancel{2} \cdot \cancel{1}} = 360$$

Although Examples 4–42 and 4–43 were solved by the multiplication rule, they can now be solved by the permutation rule.

In Example 4–42, five locations were taken and then arranged in order; hence,

$${}_5P_5 = \frac{5!}{(5-5)!} = \frac{5!}{0!} = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{1} = 120$$

(Recall that $0! = 1$.)

In Example 4–43, three locations were selected from 5 locations, so $n = 5$ and $r = 3$; hence,

$${}_5P_3 = \frac{5!}{(5-3)!} = \frac{5!}{2!} = \frac{5 \cdot 4 \cdot 3 \cdot \cancel{2} \cdot \cancel{1}}{\cancel{2} \cdot \cancel{1}} = 60$$

Examples 4–44 and 4–45 illustrate the permutation rule.

EXAMPLE 4–44 Radio Show Guests

A radio talk show host can select 3 of 6 special guests for her program. The order of appearance of the guests is important. How many different ways can this be done?

SOLUTION

Since the order of appearance on the show is important, there are ${}_6P_3$ ways to select the guests.

$${}_6P_3 = \frac{6!}{(6-3)!} = \frac{6!}{3!} = \frac{6 \cdot 5 \cdot 4 \cdot 3!}{3!} = 120$$

Hence, there can be 120 different ways to select 3 guests and present them on the program in a specific order.

EXAMPLE 4–45 School Musical Plays

A school musical director can select 2 musical plays to present next year. One will be presented in the fall, and one will be presented in the spring. If she has 9 to pick from, how many different possibilities are there?

SOLUTION

Order is important since one play can be presented in the fall and the other play in the spring.

$${}_9P_2 = \frac{9!}{(9-2)!} = \frac{9!}{7!} = \frac{9 \cdot 8 \cdot 7!}{7!} = 72$$

There are 72 different possibilities.

In the previous examples, all items involving permutations were different, but when some of the items are identical, a second permutation rule can be used.

Permutation Rule 2

The number of permutations of n objects when r_1 objects are identical, r_2 objects are identical, . . . , r_p objects are identical, etc., is

$$\frac{n!}{r_1!r_2! \cdots r_p!}$$

where $r_1 + r_2 + \cdots + r_p = n$.

EXAMPLE 4–46 Letter Permutations

How many permutations of the letters can be made from the word *STATISTICS*?

SOLUTION

In the word *STATISTICS*, there are 3 S's, 3 T's, 2 I's, 1 A, and 1 C.

$$\frac{10!}{3!3!2!1!1!} = \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{3 \cdot 2 \cdot 1 \cdot 3 \cdot 2 \cdot 1 \cdot 2 \cdot 1 \cdot 1 \cdot 1} = 50{,}400$$

There are 50,400 permutations that can be made from the word *STATISTICS*.

Combinations

Suppose a dress designer wishes to select two colors of material to design a new dress, and she has on hand four colors. How many different possibilities can there be in this situation?

OBJECTIVE 7

Find the number of ways that *r* objects can be selected from *n* objects without regard to order, using the combination rule.

This type of problem differs from previous ones in that the order of selection is not important. That is, if the designer selects yellow and red, this selection is the same as the selection red and yellow. This type of selection is called a *combination.* The difference between a permutation and a combination is that in a combination, the order or arrangement of the objects is not important; by contrast, order *is* important in a permutation. Example 4–46 illustrates this difference.

A selection of distinct objects without regard to order is called a **combination.**

© The McGraw Hill Companies, Inc./Evelyn Jo Hebert, photographer

The difference between a combination and a permutation can be shown using the letters A, B, C, and D. The permutations for the letters A, B, C, and D are

AB	BA	CA	DA
AC	BC	CB	DB
AD	BD	CD	DC

In permutations, AB is different from BA. But in combinations, AB is the same as BA since the order of the objects does not matter in combinations. Therefore, if duplicates are removed from a list of permutations, what is left is a list of combinations, as shown.

AB	~~BA~~	~~CA~~	~~DA~~
AC	BC	~~CB~~	~~DB~~
AD	BD	CD	~~DC~~

Hence, the combinations of A, B, C, and D are AB, AC, AD, BC, BD, and CD. (Alternatively, BA could be listed and AB crossed out, etc.) The combinations have been listed alphabetically for convenience, but this is not a requirement.

Interesting Fact

The total number of hours spent mowing lawns in the United States each year: 2,220,000,000.

Combinations are used when the order or arrangement is not important, as in the selecting process. Suppose a committee of 5 students is to be selected from 25 students. The 5 selected students represent a combination, since it does not matter who is selected first, second, etc.

Combination Rule

The number of combinations of r objects selected from n objects is denoted by ${}_nC_r$ and is given by the formula

$${}_nC_r = \frac{n!}{(n-r)!r!}$$

EXAMPLE 4–47 Combinations

How many combinations of 4 objects are there, taken 2 at a time?

SOLUTION

Since this is a combination problem, the answer is

$${}_4C_2 = \frac{4!}{(4-2)!2!} = \frac{4!}{2!2!} = \frac{\overset{2}{\cancel{4}} \cdot 3 \cdot \cancel{2!}}{\cancel{2} \cdot 1 \cdot \cancel{2!}} = 6$$

This is the same result shown on the previous page.

Notice that the expression for ${}_nC_r$ is

$$\frac{n!}{(n-r)!r!}$$

which is the formula for permutations with $r!$ in the denominator. In other words,

$${}_nC_r = \frac{{}_nP_r}{r!}$$

This $r!$ divides out the duplicates from the number of permutations. For each two letters, there are two permutations but only one combination. Hence, dividing the number of permutations by $r!$ eliminates the duplicates. This result can be verified for other values of n and r. *Note:* ${}_nC_n = 1$.

EXAMPLE 4–48 Movies at the Park

The director of Movies at the Park must select 4 movies from a total of 10 movies to show on Movie Night at the Park. How many different ways can the selections be made?

SOLUTION

$${}_{10}C_4 = \frac{10!}{(10-4)!4!} = \frac{10!}{6!4!} = \frac{10 \cdot 9 \cdot 8 \cdot 7}{4 \cdot 3 \cdot 2 \cdot 1} = 210$$

The director has 210 different ways to select four movies from 10 movies. In this case, the order in which the movies are shown is not important.

EXAMPLE 4–49 Committee Selection

In a club there are 7 women and 5 men. A committee of 3 women and 2 men is to be chosen. How many different possibilities are there?

SOLUTION

Here, you must select 3 women from 7 women, which can be done in ${}_7C_3$, or 35, ways. Next, 2 men must be selected from 5 men, which can be done in ${}_5C_2$, or 10, ways. Finally, by the fundamental counting rule, the total number of different ways is $35 \cdot 10 = 350$, since you are choosing both men and women. Using the formula gives

$${}_7C_3 \cdot {}_5C_2 = \frac{7!}{(7-3)!3!} \cdot \frac{5!}{(5-2)!2!} = 350$$

Table 4–1 summarizes the counting rules.

TABLE 4–1 Summary of Counting Rules

Rule	Definition	Formula
Fundamental counting rule	The number of ways a sequence of n events can occur if the first event can occur in k_1 ways, the second event can occur in k_2 ways, etc.	$k_1 \cdot k_2 \cdot k_3 \cdots k_n$
Permutation rule 1	The number of permutations of n objects taking r objects at a time (order is important)	${}_nP_r = \frac{n!}{(n-r)!}$
Permutation rule 2	The number of permutations of n objects when r_1 objects are identical, r_2 objects are identical, . . . , r_p objects are identical	$\frac{n!}{r_1!\, r_2! \cdots r_p!}$
Combination rule	The number of combinations of r objects taken from n objects (order is not important)	${}_nC_r = \frac{n!}{(n-r)!r!}$

Applying the Concepts 4–4

Garage Door Openers

Garage door openers originally had a series of four on/off switches so that homeowners could personalize the frequencies that opened their garage doors. If all garage door openers were set at the same frequency, anyone with a garage door opener could open anyone else's garage door.

1. Use a tree diagram to show how many different positions 4 consecutive on/off switches could be in.

After garage door openers became more popular, another set of 4 on/off switches was added to the systems.

2. Find a pattern of how many different positions are possible with the addition of each on/off switch.
3. How many different positions are possible with 8 consecutive on/off switches?
4. Is it reasonable to assume, if you owned a garage door opener with 8 switches, that someone could use his or her garage door opener to open your garage door by trying all the different possible positions?

For a specific year it was reported that the ignition keys for Dodge Caravans were made from a single blank that had five cuts on it. Each cut was made at one out of five possible levels. For that year assume there were 420,000 Dodge Caravans sold in the United States.

5. How many different possible keys can be made from the same key blank?
6. How many different Dodge Caravans could any one key start?

Look at the ignition key for your car and count the number of cuts on it. Assume that the cuts are made at one of any of five possible levels. Most car companies use one key blank for all their makes and models of cars.

7. Conjecture how many cars your car company sold over recent years, and then figure out how many other cars your car key could start. What would you do to decrease the odds of someone being able to open another vehicle with his or her key?

See pages 254–255 for the answers.

Exercises 4–4

1. **Zip Codes** How many 5-digit zip codes are possible if digits can be repeated? If there cannot be repetitions?

2. **Letter Permutations** List all the permutations of the letters in the word *MATH*.

3. **Speaking Order** Seven elementary students are selected to give a 3-minute presentation on what they did during summer vacation. How many different ways can the speakers be arranged?

4. **Visiting Nurses** How many different ways can a visiting nurse visit 9 patients if she wants to visit them all in one day?

5. **Quinto Lottery** A lottery game called Quinto is played by choosing five numbers each, from 0 through 9. How many numbers are possible? Although repeats are allowed, how many numbers are possible if repeats are not allowed?

6. **Show Programs** Three bands and two comics are performing for a student talent show. How many different programs (in terms of order) can be arranged? How many if the comics must perform between bands?

7. **Rolling Dice** If five dice are rolled, how many different outcomes are there?

8. **Radio Station Call Letters** The call letters of a radio station must have 4 letters. The first letter must be a K or a W. How many different station call letters can be made if repetitions are not allowed? If repetitions are allowed?

9. **Film Showings** At the Rogue Film Festival, the director must select one film from each category. There are 8 drama films, 3 sci-fi films, and 5 comedy films. How many different ways can a film be selected?

10. **Secret Code Word** How many 4-letter code words can be made using the letters in the word *pencil* if repetitions are permitted? If repetitions are not permitted?

11. **Passwords** Given the characters *A, B, C, H, I, T, U, V,* 1, 2, 3, and 4, how many seven-character passwords can be made? (No repeats are allowed.) How many if you have to use all four numbers as the first four characters in the password?

12. **Automobile Trips** There are 2 major roads from city *X* to city *Y* and 4 major roads from city *Y* to city *Z*. How many different trips can be made from city *X* to city *Z*, passing through city *Y*?

13. Evaluate each expression.

    a. 11! *b.* 9! *c.* 0! *d.* 1!
    e. ${}_6P_4$ *f.* ${}_{12}P_8$ *g.* ${}_7P_7$ *h.* ${}_4P_0$
    i. ${}_9P_2$ *j.* ${}_{11}P_3$

14. Evaluate each expression.

    a. 6! *b.* 11! *c.* 2! *d.* 9!
    e. ${}_9P_6$ *f.* ${}_{11}P_4$ *g.* ${}_8P_0$ *h.* ${}_{10}P_2$

15. **Sports Car Stripes** How many different 4-color code stripes can be made on a sports car if each code consists of the colors green, red, blue, and white? All colors are used only once.

16. **Manufacturing Tests** An inspector must select 3 tests to perform in a certain order on a manufactured part. He has a choice of 7 tests. How many ways can he perform 3 different tests?

17. **Endangered Amphibians** There are 9 endangered amphibian species in the United States. How many ways

can a student select 3 of these species to write a report about them? The order of selection is important.

18. **Inspecting Restaurants** How many different ways can a city health department inspector visit 5 restaurants in a city with 10 restaurants?

19. **Word Permutation** How many different 4-letter permutations can be written from the word *hexagon*?

20. **Cell Phone Models** A particular cell phone company offers 4 models of phones, each in 6 different colors and each available with any one of 5 calling plans. How many combinations are possible?

21. **ID Cards** How many different ID cards can be made if there are 6 digits on a card and no digit can be used more than once?

22. **Free-Sample Requests** An online coupon service has 13 offers for free samples. How many different requests are possible if a customer must request exactly 3 free samples? How many are possible if the customer may request up to 3 free samples?

23. **Ticket Selection** How many different ways can 4 tickets be selected from 50 tickets if each ticket wins a different prize?

24. **Movie Selections** The Foreign Language Club is showing a four-movie marathon of subtitled movies. How many ways can they choose 4 from the 11 available?

25. **Task Assignments** How many ways can an adviser choose 4 students from a class of 12 if they are all assigned the same task? How many ways can the students be chosen if they are each given a different task?

26. **Agency Cases** An investigative agency has 7 cases and 5 agents. How many different ways can the cases be assigned if only 1 case is assigned to each agent?

27. **Signal Flags** How many different flag signals, each consisting of 7 flags hung vertically, can be made when there are 3 indistinguishable red flags, 2 blue flags, and 2 white flags?

28. **Word Permutations** How many permutations can be made using all the letters in the word *MASSACHUSETTS*?

29. **Code Words** How many different 9-letter code words can be made using the symbols %, %, %, %, &, &, &, +, +?

30. **Toothpaste Display** How many different ways can 5 identical tubes of tartar control toothpaste, 3 identical tubes of bright white toothpaste, and 4 identical tubes of mint toothpaste be arranged in a grocery store counter display?

31. **Book Arrangements** How many different ways can 6 identical hardback books, 3 identical paperback books, and 3 identical boxed books be arranged on a shelf in a bookstore?

32. **Letter Permutations** How many different permutations of the letters in the word *CINCINNATI* are there?

33. Evaluate each expression.
    a. ${}_5C_2$
    b. ${}_8C_3$
    c. ${}_7C_4$
    d. ${}_6C_2$
    e. ${}_6C_4$

34. Evaluate each expression.
    a. ${}_3C_0$
    b. ${}_3C_3$
    c. ${}_9C_7$
    d. ${}_{12}C_2$
    e. ${}_4C_3$

35. **Medications for Depression** A researcher wishes her patients to try a new medicine for depression. How many different ways can she select 5 patients from 50 patients?

36. **Selecting Players** How many ways can 4 baseball players and 3 basketball players be selected from 12 baseball players and 9 basketball players?

37. **Coffee Selection** A coffee shop serves 12 different kinds of coffee drinks. How many ways can 4 different coffee drinks be selected?

38. **Selecting Christmas Presents** If a person can select 3 presents from 10 presents under a Christmas tree, how many different combinations are there?

39. **Buffet Desserts** In how many ways can you choose 3 kinds of ice cream and 2 toppings from a dessert buffet with 10 kinds of ice cream and 6 kinds of toppings?

40. **Bridge Foursomes** How many different tables of 4 can you make from 16 potential bridge players? How many different tables if 4 of the players insist on playing together?

41. **Music Recital** Six students are performing one song each in a jazz vocal recital. Two students have repertoires of five numbers, and the others have four songs each prepared. How many different programs are possible without regard to order? Assume that the repertory selections are all unique.

42. **Freight Train Cars** In a train yard there are 4 tank cars, 12 boxcars, and 7 flatcars. How many ways can a train be made up consisting of 2 tank cars, 5 boxcars, and 3 flatcars? (In this case, order is not important.)

43. **Selecting a Committee** There are 7 women and 5 men in a department. How many ways can a committee of 4 people be selected? How many ways can this committee be selected if there must be 2 men and 2 women on the committee? How many ways can this committee be selected if there must be at least 2 women on the committee?

44. **Selecting Cereal Boxes** Wake Up cereal comes in 2 types, crispy and crunchy. If a researcher has 10 boxes of each, how many ways can she select 3 boxes of each for a quality control test?

45. **Hawaiian Words** The Hawaiian alphabet consists of 7 consonants and 5 vowels. How many three-letter "words" are possible if there are never two consonants together and if a word must always end in a vowel?

46. **Selecting a Jury** How many ways can a jury of 6 women and 6 men be selected from 10 women and 12 men?

47. **Selecting Students** How many ways can you pick 4 students from 10 students (6 men, 4 women) if you must have an equal number of each gender or all of the same gender?

48. **Investigative Team** The state narcotics bureau must form a 5-member investigative team. If it has 25 agents from which to choose, how many different possible teams can be formed?

49. **Dominoes** A domino is a flat rectangular block whose face is divided into two square parts, each part showing from zero to six pips (or dots). Playing a game consists of playing dominoes with a matching number of pips. Explain why there are 28 dominoes in a complete set.

50. **Charity Event Participants** There are 16 seniors and 15 juniors in a particular social organization. In how many ways can 4 seniors and 2 juniors be chosen to participate in a charity event?

51. **Automobile Selection** An automobile dealer has 12 small automobiles, 8 mid-size automobiles, and 6 large automobiles on his lot. How many ways can two of each type of automobile be selected from his inventory?

52. **DVD Selection** How many ways can a person select 8 DVDs from a display of 13 DVDs?

53. **Railroad Accidents** A researcher wishes to study railroad accidents. He wishes to select 3 railroads from 10 Class I railroads, 2 railroads from 6 Class II railroads, and 1 railroad from 5 Class III railroads. How many different possibilities are there for his study?

54. **Selecting a Location** An advertising manager decides to have an ad campaign in which 8 special calculators will be hidden at various locations in a shopping mall. If he has 17 locations from which to pick, how many different possible combinations can he choose?

Permutations and Combinations

55. **Selecting Posters** A buyer decides to stock 8 different posters. How many ways can she select these 8 if there are 20 from which to choose?

56. **Test Marketing Products** Anderson Research Company decides to test-market a product in 6 areas. How many different ways can 3 areas be selected in a certain order for the first test?

57. **Nuclear Power Plants** How many different ways can a government researcher select 5 nuclear power plants from 9 nuclear power plants in Pennsylvania?

58. **Selecting Musicals** How many different ways can a theatrical group select 2 musicals and 3 dramas from 11 musicals and 8 dramas to be presented during the year?

59. **Textbook Selection** How many different ways can an instructor select 2 textbooks from a possible 17?

60. **DVD Selection** How many ways can a person select 8 DVDs from 10 DVDs?

61. **Flight Attendants** How many different ways can 3 flight attendants be selected from 11 flight attendants for a routine flight?

62. **Signal Flags** How many different signals can be made by using at least 3 different flags if there are 5 different flags from which to select?

63. **Dinner Selections** How many ways can a dinner patron select 3 appetizers and 2 vegetables if there are 6 appetizers and 5 vegetables on the menu?

64. **Air Pollution** The Environmental Protection Agency must investigate 9 mills for complaints of air pollution. How many different ways can a representative select 5 of these to investigate this week?

65. **Selecting Officers** In a board of directors composed of 8 people, how many ways can one chief executive officer, one director, and one treasurer be selected?

66. **Selecting Council Members** The presidents, vice presidents, and secretary-treasurers from each of four classes are eligible for an all-school council. How many ways can four officers be chosen from these representatives? How many ways can they be chosen if the president must be selected from the sitting presidents, the vice president from the sitting vice presidents, the secretary from the sitting secretary-treasurers, and the treasurer from everybody who's left?

Extending the Concepts

67. **Selecting Coins** How many different ways can you select one or more coins if you have 2 nickels, 1 dime, and 1 half-dollar?

68. **People Seated in a Circle** In how many ways can 3 people be seated in a circle? 4? n? (*Hint:* Think of them standing in a line before they sit down and/or draw diagrams.)

69. **Seating in a Movie Theater** How many different ways can 5 people—A, B, C, D, and E—sit in a row at a movie theater if (*a*) A and B must sit together; (*b*) C must sit to the right of, but not necessarily next to, B; (*c*) D and E will not sit next to each other?

70. **Poker Hands** Using combinations, calculate the number of each type of poker hand in a deck of

cards. (A poker hand consists of 5 cards dealt in any order.)

a. Royal flush

b. Straight flush (not including a royal flush)

c. Four of a kind

d. Full house

71. How many different combinations can be made from $(x + 2)$ things taken x at a time?

72. A game of concentration (memory) is played with a standard 52-card deck. How many potential two-card matches are there (e.g., one jack "matches" any other jack)?

Technology Step by Step

TI-84 Plus Step by Step

Factorials, Permutations, and Combinations

Factorials *n*!

1. Type the value of n.
2. Press **MATH** and move the cursor to PRB, then press **4** for !.
3. Press **ENTER.**

Permutations $_nP_r$

1. Type the value of n.
2. Press **MATH** and move the cursor to PRB, then press **2** for $_nP_r$.
3. Type the value of r.
4. Press **ENTER.**

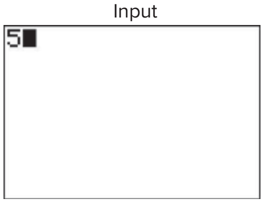
Combinations $_nC_r$

1. Type the value of n.
2. Press **MATH** and move the cursor to PRB, then press **3** for $_nC_r$.
3. Type the value of r.
4. Press **ENTER.**

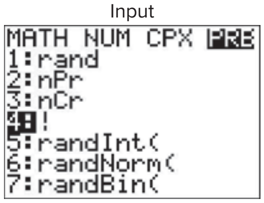
Example TI 4–2

Calculate 5! (Example 4–42 from the text).

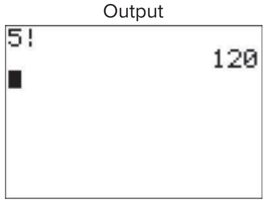
Input



Input

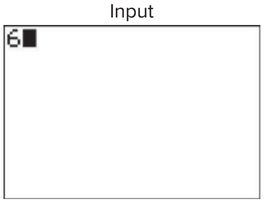


Output

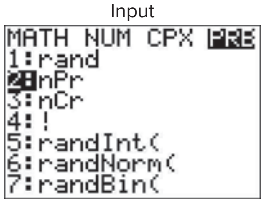


Calculate $_6P_3$ (Example 4–44 from the text).

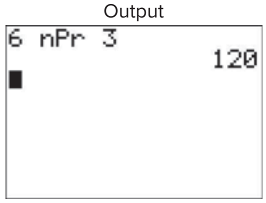
Input



Input



Output



Calculate ${}_{10}C_3$ (Example 4–48 from the text).

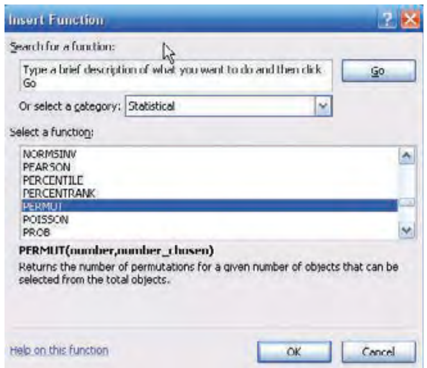
Input Input Output

EXCEL Step by Step

Permutations, Combinations, and Factorials

To find a value of a permutation, for example, ${}_5P_3$:

1. In an open cell in an Excel worksheet, select the Formulas tab on the toolbar. Then click the Insert function icon.
2. Select the Statistical function category, then the PERMUT function, and click [OK].

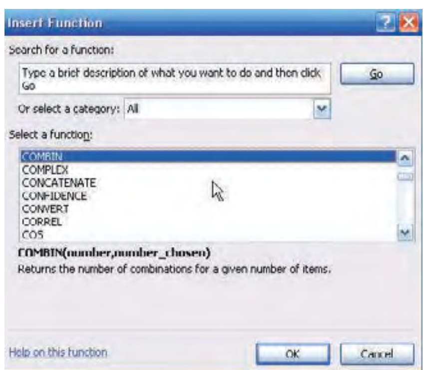


3. Type **5** in the Number box.
4. Type **3** in the Number_chosen box and click [OK].

The selected cell will display the answer: 60.

To find a value of a combination, for example, ${}_5C_3$:

1. In an open cell, select the Formulas tab on the toolbar. Click the Insert function icon.
2. Select the All function category, then the COMBIN function, and click [OK].



3. Type **5** in the Number box.
4. Type **3** in the Number_chosen box and click [OK].

The selected cell will display the answer: 10.

To find a factorial of a number, for example, 7!:

1. In an open cell, select the Formulas tab on the toolbar. Click the Insert function icon.
2. Select the Math & Trig function category, then the FACT function, and click [OK].

3. Type **7** in the Number box and click [OK].

The selected cell will display the answer: 5040.

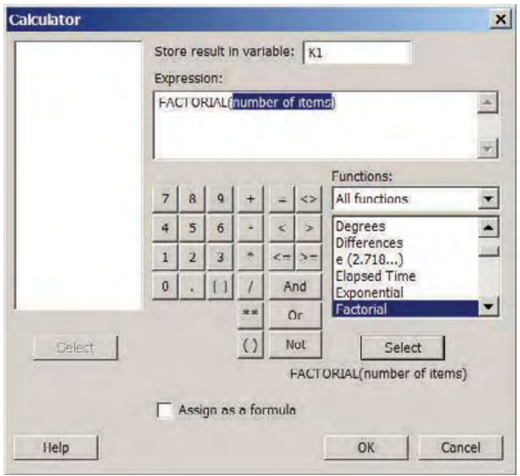
MINITAB Step by Step

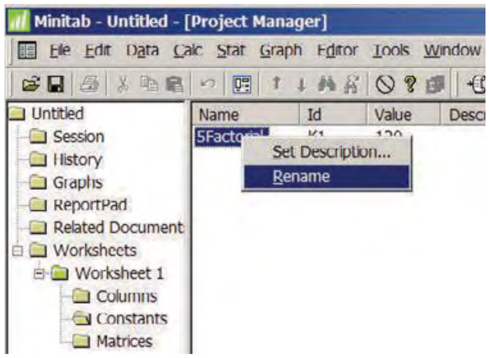
Factorials, Permutations, and Combinations

We will use Minitab to calculate 5!, ${}_6P_3$, and ${}_{10}C_3$

The results are stored in a constant; MINITAB has 1000 of them available in each worksheet. They are numbered from K1 to K1000. By default, MINITAB assigns the values of missing constants, *e*, and pi to the last three stored constants: K998 = *; K999 = 2.71828; and K1000 = 3.14159. To reduce clutter, MINITAB does not list these in the Constants subfolder.

1. To calculate 5!, select **Calc>Calculator.**
   a) Type K1 in the Store result in variable, then press the Tab key on your keyboard or click in the Expression dialog box.
   b) Scroll down the function list to Factorial, then click [Select].
   c) Type a 5 to replace the number of items that is highlighted.
   d) Click [OK]. The value of 5! = 120 will be stored in a constant that is not visible.
   e) To see the constant, click the icon for Project Manager, then the Constants item in the worksheet folder. Right-click on the Untitled name, then choose Rename and type 5Factorial. Storing the constant makes it available to be used in future calculations.



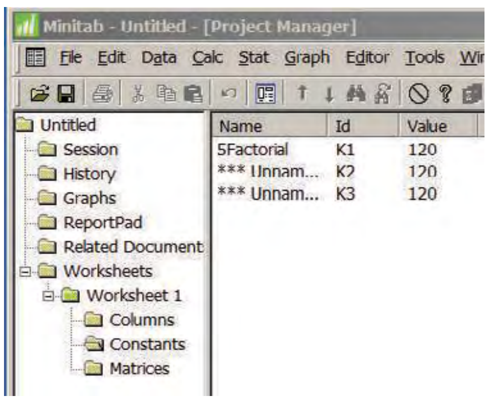


f) To view the constants in the Session window, select **Data>Display Data** then choose K1 5factorial and [OK].

2. To calculate ${}_6P_3$, select **Calc>Calculator**.
   a) Type K2 in the Store result in variable, then press the Tab key on your keyboard or click in the Expression dialog box.
   b) Scroll down the function list to Permutations, then click [Select].
   c) Type a 6 to replace the number of items, then 3 to replace number to choose.
   d) Click [OK]. Name the constant and display it, using the instructions in steps 1e and 1f.

3. To calculate ${}_{10}C_3$, select **Calc>Calculator**.
   a) Type K3 in the Store result in variable, then press the Tab key on your keyboard or click in the Expression dialog box.
   b) Scroll down the function list to Combinations, then click [Select].
   c) Type a 10 to replace the number of items, then 3 to replace number to choose.
   d) Click [OK]. Name the constant and display it, using the instructions in steps 1e and 1f.

The Project Manager is shown. Coincidence! The values are all 120. These values can be very large and exceed the storage capability for constants. Values larger in absolute value than 1.001000E+18 in K1 are converted to missing. In scientific notation that is 1.001000×10^{18}.



4–5 Probability and Counting Rules

OBJECTIVE 8

Find the probability of an event, using the counting rules.

The counting rules can be combined with the probability rules in this chapter to solve many types of probability problems. By using the fundamental counting rule, the permutation rules, and the combination rule, you can compute the probability of outcomes of many experiments, such as getting a full house when 5 cards are dealt or selecting a committee of 3 women and 2 men from a club consisting of 10 women and 10 men.

EXAMPLE 4–50 Four Aces

Find the probability of getting 4 aces when 5 cards are drawn from an ordinary deck of cards.

SOLUTION

There are ${}_{52}C_5$ ways to draw 5 cards from a deck. There is only 1 way to get 4 aces (that is, ${}_4C_4$), but there are 48 possibilities to get the fifth card. Therefore, there are 48 ways to get 4 aces and 1 other card. Hence,

$$P(4\text{ aces}) = \frac{{}_4C_4 \cdot 48}{{}_{52}C_5} = \frac{1 \cdot 48}{2{,}598{,}960} = \frac{48}{2{,}598{,}960} = \frac{1}{54{,}145}$$

EXAMPLE 4–51 Defective Integrated Circuits

A box contains 24 integrated circuits, 4 of which are defective. If 4 are sold at random, find the following probabilities.

a. Exactly 2 are defective.

b. None is defective.

c. All are defective.

d. At least 1 is defective.

SOLUTION

There are ${}_{24}C_4$ ways to sell 4 integrated circuits, so the denominator in each case will be 10,626.

a. Two defective integrated circuits can be selected as ${}_4C_2$ and two nondefective ones as ${}_{20}C_2$. Hence,

$$P(\text{exactly 2 defectives}) = \frac{{}_4C_2 \cdot {}_{20}C_2}{{}_{24}C_4} = \frac{1140}{10{,}626} = \frac{190}{1771}$$

b. The number of ways to choose no defectives is ${}_{20}C_4$. Hence,

$$P(\text{no defectives}) = \frac{{}_{20}C_4}{{}_{24}C_4} = \frac{4845}{10{,}626} = \frac{1615}{3542}$$

c. The number of ways to choose 4 defectives from 4 is ${}_4C_4$, or 1. Hence,

$$P(\text{all defective}) = \frac{1}{{}_{24}C_4} = \frac{1}{10{,}626}$$

d. To find the probability of at least 1 defective transistor, find the probability that there are no defective integrated circuits, and then subtract that probability from 1.

$$\begin{aligned} P(\text{at least 1 defective}) &= 1 - P\,(\text{no defectives}) \\ &= 1 - \frac{{}_{20}C_4}{{}_{24}C_4} = 1 - \frac{1615}{3542} = \frac{1927}{3542} \end{aligned}$$

EXAMPLE 4–52 Term Paper Selection

A student needs to select two topics to write two term papers for a course. There are 8 topics in economics and 11 topics in science. Find the probability that she selects one topic in economics and one topic is science to complete her assignment.

SOLUTION

$$\begin{aligned} P(\text{economics and science}) &= \frac{{}_8C_1 \cdot {}_{11}C_1}{{}_{19}C_2} \\ &= \frac{88}{171} = 0.515 \approx 51.5\% \end{aligned}$$

Hence, there is about a 51.5% probability that a person will select one topic from economics and one topic from science.

EXAMPLE 4–53 State Lottery Number

In the Pennsylvania State Lottery, a person selects a three-digit number and repetitions are permitted. If a winning number is selected, find the probability that it will have all three digits the same.

SOLUTION

Since there are 10 different digits, there are $10 \cdot 10 \cdot 10 = 1000$ ways to select a winning number. When all of the digits are the same, that is, 000, 111, 222, . . . , 999, there are 10 possibilities, so the probability of selecting a winning number which has 3 identical digits is $\frac{10}{1000} = \frac{1}{100}$.

EXAMPLE 4–54 Tennis Tournament

There are 8 married couples in a tennis club. If 1 man and 1 woman are selected at random to plan the summer tournament, find the probability that they are married to each other.

SOLUTION

Since there are 8 ways to select the man and 8 ways to select the woman, there are $8 \cdot 8$, or 64, ways to select 1 man and 1 woman. Since there are 8 married couples, the solution is $\frac{8}{64} = \frac{1}{8}$.

As indicated at the beginning of this section, the counting rules and the probability rules can be used to solve a large variety of probability problems found in business, gambling, economics, biology, and other fields.

SPEAKING OF STATISTICS The Mathematics of Gambling

Gambling is big business. There are state lotteries, casinos, sports betting, and church bingos. It seems that today everybody is either watching or playing Texas Hold 'Em Poker.

Using permutations, combinations, and the probability rules, mathematicians can find the probabilities of various gambling games. Here are the probabilities of the various 5-card poker hands.

Hand	Number of ways	Probability
Straight flush	40	0.000015
Four of a kind	624	0.000240
Full house	3,744	0.001441
Flush	5,108	0.001965
Straight	10,200	0.003925
Three of a kind	54,912	0.021129
Two pairs	123,552	0.047539
One pair	1,098,240	0.422569
Less than one pair	1,302,540	0.501177
Total	2,598,960	1.000000

© Royalty-Free/CORBIS

The chance of winning at gambling games can be compared by using what is called the house advantage, house edge, or house percentage. For example, the house advantage for roulette is about 5.26%, which means in the long run, the house wins 5.26 cents on every $1 bet; or you will lose, on average, 5.26 cents on every $1 you bet. The lower the house advantage, the more favorable the game is to you.

For the game of craps, the house advantage is anywhere between 1.4 and 15%, depending on what you bet on. For the game called Keno, the house advantage is 29.5%. The house advantage for Chuck-a-Luck is 7.87%, and for Baccarat, it is either 1.36 or 1.17%, depending on your bet.

Slot machines have a house advantage anywhere from about 4 to 10% depending on the geographic location, such as Atlantic City, Las Vegas, and Mississippi, and the amount put in the machine, such as 5 cents, 25 cents, and $1.

Actually, gamblers found winning strategies for the game blackjack or 21, such as card counting. However, the casinos retaliated by using multiple decks and by banning card counters.

Applying the Concepts 4–5

Counting Rules and Probability

One of the biggest problems for students when doing probability problems is to decide which formula or formulas to use. Another problem is to decide whether two events are independent or dependent. Use the following problem to help develop a better understanding of these concepts.

Assume you are given a five-question multiple-choice quiz. Each question has 5 possible answers: A, B, C, D, and E.

1. How many events are there?
2. Are the events independent or dependent?
3. If you guess at each question, what is the probability that you get all of them correct?
4. What is the probability that a person guesses answer A for each question?

Assume that you are given a five-question matching test in which you are to match the correct answers in the right column with the questions in the left column. You can use each answer only once.

5. How many events are there?
6. Are the events independent or dependent?
7. What is the probability of getting them all correct if you are guessing?
8. What is the difference between the two problems?

See pages 254–255 for the answers.

Exercises 4–5

1. **Selecting Cards** Find the probability of getting 2 face cards (king, queen, or jack) when 2 cards are drawn from a deck without replacement.

2. **Selecting Cards** Cards numbered 1–10 are shuffled and dealt face down. What is the probability that they are in order?

3. **Educational Fellowship** A university received 9 applications for three postdoctorate fellowships. Five of the applicants are men and four are women. Find these probabilities.
   a. All 3 who are selected are men.
   b. All 3 who are selected are women.
   c. Two men and one woman are selected.
   d. Two women and one man are selected.

4. **Senate Partisanship** The composition of the Senate of the 114th Congress is

   54 Republicans   2 Independent   44 Democrats

   A new committee is being formed to study ways to benefit the arts in education. If 3 Senators are selected at random to form a new committee, what is the probability that they will all be Republicans? What is the probability that they will all be Democrats? What is the probability that there will be 1 from each party, including the Independent?

   Source: New York Times Almanac.

5. **Job Applications** Six men and seven women apply for two identical jobs. If the jobs are filled at random, find the following:
   a. The probability that both are filled by men.
   b. The probability that both are filled by women.
   c. The probability that one man and one woman are hired.
   d. The probability that the one man and one woman who are twins are hired.

6. **Defective Resistors** A package contains 12 resistors, 3 of which are defective. If 4 are selected, find the probability of getting
   a. 0 defective resistors
   b. 1 defective resistor
   c. 3 defective resistors

7. **Winning Tickets** At a meeting of 10 executives (7 women and 3 men), two door prizes are awarded. Find the probability that both prizes are won by men.

8. **Getting a Full House** Find the probability of getting a full house (3 cards of one denomination and 2 of another) when 5 cards are dealt from an ordinary deck.

9. **World-Class Orchestras** About.com's list of 20 World Class Orchestras includes the following from the United States: Boston Symphony Orchestra, Chicago Symphony Orchestra, Cleveland Orchestra, Los Angeles Philharmonic, New York Philharmonic, the Metropolitan Opera Orchestra, and the San Francisco Symphony. Choose 5 at random from the list of 20 for a benefit CD. What is the probability that the collection will include at least one group from the United States? At least 2 from the United States? That all 5 will be from the United States?

10. **Selecting Cards** The red face cards and the black cards numbered 2–9 are put into a bag. Four cards are drawn at random without replacement. Find the following probabilities.
    a. All 4 cards are red.
    b. 2 cards are red and 2 cards are black.
    c. At least 1 of the cards is red.
    d. All 4 cards are black.

11. **Socks in a Drawer** A drawer contains 11 identical red socks and 8 identical black socks. Suppose that you choose 2 socks at random in the dark.
    a. What is the probability that you get a pair of red socks?
    b. What is the probability that you get a pair of black socks?
    c. What is the probability that you get 2 unmatched socks?
    d. Where did the other red sock go?

12. **Selecting Books** Find the probability of selecting 3 science books and 4 math books from 8 science books and 9 math books. The books are selected at random.

13. **Rolling the Dice** If three dice are rolled, find the probability of getting a sum of 6.

14. **Football Team Selection** A football team consists of 20 freshmen and 20 sophomores, 15 juniors, and 10 seniors. Four players are selected at random to serve as captains. Find the probability that
   a. All 4 are seniors
   b. There is 1 each: freshman, sophomore, junior, and senior
   c. There are 2 sophomores and 2 freshmen
   d. At least 1 of the students is a senior

15. **Arrangement of Washers** Find the probability that if 5 different-sized washers are arranged in a row, they will be arranged in order of size.

16. **Poker Hands** Using the information in Exercise 70 in Section 4–4, find the probability of each poker hand.
   a. Royal flush
   b. Straight flush
   c. Four of a kind

17. **Plant Selection** All holly plants are dioecious—a male plant must be planted within 30 to 40 feet of the female plants in order to yield berries. A home improvement store has 12 unmarked holly plants for sale, 8 of which are female. If a homeowner buys 3 plants at random, what is the probability that berries will be produced?

Summary

In this chapter, the basic concepts of probability are explained.

- There are three basic types of probability: classical probability, empirical probability, and subjective probability. Classical probability uses sample spaces. Empirical probability uses frequency distributions, and subjective probability uses an educated guess to determine the probability of an event. The probability of any event is a number from 0 to 1. If an event cannot occur, the probability is 0. If an event is certain, the probability is 1. The sum of the probability of all the events in the sample space is 1. To find the probability of the complement of an event, subtract the probability of the event from 1. (4–1)
- Two events are mutually exclusive if they cannot occur at the same time; otherwise, the events are not mutually exclusive. To find the probability of two mutually exclusive events occurring, add the probability of each event. To find the probability of two events when they are not mutually exclusive, add the possibilities of the individual events and then subtract the probability that both events occur at the same time. These types of probability problems can be solved by using the addition rules. (4–2)
- Two events are independent if the occurrence of the first event does not change the probability of the second event occurring. Otherwise, the events are dependent. To find the probability of two independent events occurring, multiply the probabilities of each event. To find the probability that two dependent events occur, multiply the probability that the first event occurs by the probability that the second event occurs, given that the first event has already occurred. The complement of an event is found by selecting the outcomes in the sample space that are not involved in the outcomes of the event. These types of problems can be solved by using the multiplication rules and the complementary event rules. (4–3)
- Finally, when a large number of events can occur, the fundamental counting rule, the permutation rules, and the combination rule can be used to determine the number of ways that these events can occur. (4–4)
- The counting rules and the probability rules can be used to solve more-complex probability problems. (4–5)

Important Terms

classical probability 189
combination 232
complement of an event 192
compound event 189
conditional probability 215
dependent events 215
disjoint events 202
empirical probability 194
equally likely events 189
event 188
fundamental counting rule 226
independent events 213
law of large numbers 196
mutually exclusive events 202
outcome 186
permutation 229
probability 186
probability experiment 186
sample space 186
simple event 189
subjective probability 196
tree diagram 188
Venn diagrams 193

Important Formulas

Formula for classical probability:

$$P(E) = \frac{\text{number of outcomes in } E}{\text{total number of outcomes in sample space}} = \frac{n(E)}{n(S)}$$

Formula for empirical probability:

$$P(E) = \frac{\text{frequency for class}}{\text{total frequencies in distribution}} = \frac{f}{n}$$

Addition rule 1, for two mutually exclusive events:

$$P(A \text{ or } B) = P(A) + P(B)$$

Addition rule 2, for events that are not mutually exclusive:

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

Multiplication rule 1, for independent events:

$$P(A \text{ and } B) = P(A) \cdot P(B)$$

Multiplication rule 2, for dependent events:

$$P(A \text{ and } B) = P(A) \cdot P(B|A)$$

Formula for conditional probability:

$$P(B|A) = \frac{P(A \text{ and } B)}{P(A)}$$

Formula for complementary events:

$$P(\overline{E}) = 1 - P(E) \quad \text{or} \quad P(E) = 1 - P(\overline{E})$$

$$\text{or} \quad P(E) + P(\overline{E}) = 1$$

Fundamental counting rule: In a sequence of n events in which the first one has k_1 possibilities, the second event has k_2 possibilities, the third has k_3 possibilities, etc., the total number of possibilities of the sequence will be

$$k_1 \cdot k_2 \cdot k_3 \cdots k_n$$

Permutation rule 1: The number of permutations of n objects taking r objects at a time when order is important is

$${}_nP_r = \frac{n!}{(n-r)!}$$

Permutation rule 2: The number of permutations of n objects when r_1 objects are identical, r_2 objects are identical, . . . , r_p objects are identical is

$$\frac{n!}{r_1!\, r_1! \cdots r_p!}$$

Combination rule: The number of combinations of r objects selected from n objects when order is not important is

$${}_nC_r = \frac{n!}{(n-r)!r!}$$

Review Exercises

Section 4–1

1. **Rolling a Die** An eight-sided die is rolled. Find the probability of each.
   a. Getting a 6
   b. Getting a number larger than 5
   c. Getting an odd number

2. **Selecting a Card** When a card is selected from a deck, find the probability of getting
   a. A club
   b. A face card or a heart
   c. A 6 and a spade
   d. A king
   e. A red card

3. **Software Selection** The top-10 selling computer software titles in a recent year consisted of 3 for doing taxes, 5 antivirus or security programs, and 2 “other.” Choose one title at random.
   a. What is the probability that it is not used for doing taxes?
   b. What is the probability that it is used for taxes or is one of the “other” programs?

   Source: www.infoplease.com

4. **Motor Vehicle Producers** The top five motor vehicle producers in the world are listed below with the number of vehicles produced in 2010 (in thousands of vehicles).

China	16,144
Japan	9,197
United States	7,632
Germany	5,700
South Korea	4,184

   Choose one vehicle at random;
   a. What is the probability that it was produced in the United States?
   b. What is the probability that it was not produced in Asia?
   c. What is the probability that it was produced in Germany or Japan?

   Source: World Almanac 2012.

5. **Exercise Preference** In a local fitness club, 32% of the members lifted weights to exercise, 41% preferred aerobics, and 6% preferred both. If a member is selected at random, find the probability that the member prefers another method (neither weights nor aerobics) to exercise.

6. **Rolling Two Dice** When two dice are rolled, find the probability of getting
   a. A sum of 5 or 6
   b. A sum greater than 9
   c. A sum less than 4 or greater than 9
   d. A sum that is divisible by 4
   e. A sum of 14
   f. A sum less than 13

Section 4–2

7. **New Cars** The probability that a new automobile has a backup camera is 0.6. The probability that a new automobile has a GPS system is 0.4. The probability that a new automobile has both a backup camera and a GPS is 0.2. If a new automobile is selected at random, find the probability that it has neither a backup camera nor a GPS.

8. **Breakfast Drink** In a recent survey, 18 people preferred milk, 29 people preferred coffee, and 13 people preferred juice as their primary drink for breakfast. If a person is selected at random, find the probability that the person preferred juice as her or his primary drink.

9. **Lawnmower and Weed Wacker Ownership** The probability that a homeowner owns a lawnmower is 0.7. The probability that a homeowner owns a weed wacker is 0.5. The probability that a home owner owns both a lawnmower and a weed wacker is 0.3. If a homeowner is selected at random, find the probability that he or she owns either a lawnmower or a weed wacker.

10. **Casino Gambling** The probability that a gambler plays table games is 0.32. The probability that a person plays the slot machines is 0.85. The probability that a person plays both is 0.15. A gambler can play more than one type of game. Find the probability that a gambler plays table games given that this person plays the slot machines.

11. **Online Course Selection** Roughly 1 in 6 students enrolled in higher education took at least one online course last fall. Choose 5 enrolled students at random. Find the probability that
    a. All 5 took online courses
    b. None of the 5 took a course online
    c. At least 1 took an online course

    Source: www.encarta.msn.com

12. **Purchasing Sweaters** During a sale at a men's store, 16 white sweaters, 3 red sweaters, 9 blue sweaters, and 7 yellow sweaters were purchased. If a customer is selected at random, find the probability that he bought
    a. A blue sweater
    b. A yellow or a white sweater
    c. A red, a blue, or a yellow sweater
    d. A sweater that was not white

Section 4–3

13. **Drawing Cards** Three cards are drawn from an ordinary deck *without* replacement. Find the probability of getting
    a. All black cards
    b. All spades
    c. All queens

14. **Coin Toss and Card Drawn** A coin is tossed and a card is drawn from a deck. Find the probability of getting
    a. A head and a 6
    b. A tail and a red card
    c. A head and a club

15. **Movie Releases** The top five countries for movie releases for a specific year are the United States with 471 releases, United Kingdom with 386, Japan with 79, Germany with 316, and France with 132. Choose 1 new release at random. Find the probability that it is
    a. European
    b. From the United States
    c. German or French
    d. German given that it is European

    Source: www.showbizdata.com

16. **Factory Output** A manufacturing company has three factories: X, Y, and Z. The daily output of each is shown here.

Product	Factory X	Factory Y	Factory Z
TVs	18	32	15
Stereos	6	20	13

    If 1 item is selected at random, find these probabilities.
    a. It was manufactured at factory X or is a stereo.
    b. It was manufactured at factory Y or factory Z.
    c. It is a TV or was manufactured at factory Z.

17. **Effectiveness of Vaccine** A vaccine has a 90% probability of being effective in preventing a certain disease. The probability of getting the disease if a person is not vaccinated is 50%. In a certain geographic region, 25% of the people get vaccinated. If a person is selected at random, find the probability that he or she will contract the disease.

18. **T-shirt Factories** Two T-shirt printing factories produce T-shirts for a local sports team. Factory A produces 60% of the shirts and factory B produces 40%. Five percent of the shirts from factory A are defective, and 6% of the shirts from factory B are defective. Choose 1 shirt at random. Given that the shirt is defective, what is the probability that it came from factory A?

19. **Car Purchase** The probability that Sue will live on campus and buy a new car is 0.37. If the probability that she will live on campus is 0.73, find the probability that she will buy a new car, given that she lives on campus.

20. **Applying Shipping Labels** Four unmarked packages have lost their shipping labels, and you must reapply them. What is the probability that you apply the labels and get all 4 of them correct? Exactly 3 correct? Exactly 2? At least 1 correct?

21. **Health Club Membership** Of the members of the Blue River Health Club, 43% have a lifetime membership and

exercise regularly (three or more times a week). If 75% of the club members exercise regularly, find the probability that a randomly selected member is a life member, given that he or she exercises regularly.

22. **Bad Weather** The probability that it snows and the bus arrives late is 0.023. José hears the weather forecast, and there is a 40% chance of snow tomorrow. Find the probability that the bus will be late, given that it snows.

23. **Education Level and Smoking** At a large factory, the employees were surveyed and classified according to their level of education and whether they smoked. The data are shown in the table.

	Educational level		
Smoking habit	**Not high school graduate**	**High school graduate**	**College graduate**
Smoke	6	14	19
Do not smoke	18	7	25

If an employee is selected at random, find these probabilities.

a. The employee smokes, given that he or she graduated from college.

b. Given that the employee did not graduate from high school, he or she is a smoker.

24. **War Veterans** Approximately 11% of the civilian population are veterans. Choose 5 civilians at random. What is the probability that none are veterans? What is the probability that at least 1 is a veteran?

Source: www.factfinder.census.gov

25. **Television Sets** If 98% of households have at least one television set and 4 households are selected, find the probability that at least one household has a television set.

26. **Chronic Sinusitis** The U.S. Department of Health and Human Services reports that 15% of Americans have chronic sinusitis. If 5 people are selected at random, find the probability that at least 1 has chronic sinusitis.

Source: 100% American.

Section 4–4

27. **Motorcycle License Plates** If a motorcycle license plate consists of two letters followed by three digits, how many different license plates can be made if repetitions are allowed? How many different license plates can be made if repetitions are not allowed? How many license plates can be made if repetitions are allowed in the digits but not in the letters?

28. **Types of Copy Paper** White copy paper is offered in 5 different strengths and 11 different degrees of brightness, recycled or not, and acid-free or not. How many different types of paper are available for order?

29. **Baseball Players** How many ways can 3 outfielders and 4 infielders be chosen from 5 outfielders and 7 infielders?

30. **Carry-on Items** The following items are allowed as airline carry-on items: (1) safety razors, (2) eyedrops and saline, (3) nail clippers and tweezers, (4) blunt-tipped scissors, (5) mobile phones, (6) umbrellas, (7) common lighters, (8) beverages purchased after security screening, and (9) musical instruments. Suppose that your airline allows only 6 of these items. In how many ways can you pick 3 not to take?

31. **Names for Boys** The top 10 names for boys in America in 2005 were Ethan, Jacob, Ryan, Matthew, Tyler, Jack, Joshua, Andrew, Noah, and Michael. The top 10 names for 2015 are Liam, Noah, Ethan, Mason, Lucas, Logan, Oliver, Jackson, Aiden, and Jacob. In how many ways can you choose 5 names from these lists?

32. **Committee Representation** There are 6 Republican, 5 Democrat, and 4 Independent candidates. How many different ways can a committee of 3 Republicans, 2 Democrats, and 1 Independent be selected?

33. **Song Selections** A promotional MP3 player is available with the capacity to store 100 songs, which can be reordered at the push of a button. How many different arrangements of these songs are possible? (*Note:* Factorials get very big, very fast! How large a factorial will your calculator calculate?)

34. **Employee Health Care Plans** A new employee has a choice of 5 health care plans, 3 retirement plans, and 2 different expense accounts. If a person selects 1 of each option, how many different options does she or he have?

35. **Course Enrollment** There are 12 students who wish to enroll in a particular course. There are only 4 seats left in the classroom. How many different ways can 4 students be selected to attend the class?

36. **Candy Selection** A candy store allows customers to select 3 different candies to be packaged and mailed. If there are 13 varieties available, how many possible selections can be made?

37. **House Numbers** A home improvement store has the following house numbers: 335666 left after a sale. How many different six-digit house numbers can be made from those numbers?

38. **Word Permutations** From which word can you make more permutations, *MATHEMATICS* or *PROBABILITY*? How many more?

39. **Movie Selections** If a person can select one movie for each week night except Saturday, how many different ways can the selection be made if the person can select from 16 movies?

40. **Course Selection** If a student can select one of 3 language courses, one of 5 mathematics courses,

and one of 4 history courses, how many different schedules can be made?

Section 4–5

41. Catalog ID Numbers In a catalog, movies are identified by an ID number—2 letters followed by 3 digits. Repetitions are permitted. How many catalog numbers can be made? What is the probability that the ID number is divisible by 5? (Include the 2 letters in the ID number.)

42. License Plates A certain state's license plate has 3 letters followed by 4 numbers. Repeats are not allowed for the letters, but they are for the numbers. How many such license plates are possible? If they are issued at random, what is the probability that the 3 letters are 3 consecutive letters in alphabetical order?

43. Territorial Selection Several territories and colonies today are still under the jurisdiction of another country. France holds the most with 16 territories, the United Kingdom has 15, the United States has 14, and several other countries have territories as well. Choose 3 territories at random from those held by France, the United Kingdom, and the United States. What is the probability that all 3 belong to the same country?

Source: www.infoplease.com

44. Yahtzee Yahtzee is a game played with 5 dice. Players attempt to score points by rolling various combinations. When all 5 dice show the same number, it is called a *Yahtzee* and scores 50 points for the first one and 100 points for each subsequent Yahtzee in the same game. What is the probability that a person throws a Yahtzee on the very first roll? What is the probability that a person throws two Yahtzees on two successive turns?

45. Personnel Classification For a survey, a subject can be classified as follows:

Gender: male or female

Marital status: single, married, widowed, divorced

Occupation: administration, faculty, staff

Draw a tree diagram for the different ways a person can be classified.

STATISTICS TODAY

Would You Bet Your Life?—Revisited

In his book *Probabilities in Everyday Life*, John D. McGervey states that the chance of being killed on any given commercial airline flight is almost 1 in 1 million and that the chance of being killed during a transcontinental auto trip is about 1 in 8000. The corresponding probabilities are $1/1{,}000{,}000 = 0.000001$ as compared to $1/8000 = 0.000125$. Since the second number is 125 times greater than the first number, you have a much higher risk driving than flying across the United States.

Chapter Quiz

Determine whether each statement is true or false. If the statement is false, explain why.

1. Subjective probability has little use in the real world.
2. Classical probability uses a frequency distribution to compute probabilities.
3. In classical probability, all outcomes in the sample space are equally likely.
4. When two events are not mutually exclusive, $P(A \text{ or } B) = P(A) + P(B)$.
5. If two events are dependent, they must have the same probability of occurring.
6. An event and its complement can occur at the same time.
7. The arrangement ABC is the same as BAC for combinations.
8. When objects are arranged in a specific order, the arrangement is called a combination.

Select the best answer.

9. The probability that an event happens is 0.42. What is the probability that the event won't happen?
   a. −0.42
   b. 0.58
   c. 0
   d. 1
10. When a meteorologist says that there is a 30% chance of showers, what type of probability is the person using?
    a. Classical
    b. Empirical
    c. Relative
    d. Subjective

11. The sample space for tossing 3 coins consists of how many outcomes?

a. 2 *c.* 6
b. 4 *d.* 8

12. The complement of guessing 5 correct answers on a 5-question true/false exam is

a. Guessing 5 incorrect answers
b. Guessing at least 1 incorrect answer
c. Guessing at least 1 correct answer
d. Guessing no incorrect answers

13. When two dice are rolled, the sample space consists of how many events?

a. 6 *c.* 36
b. 12 *d.* 54

14. What is $_nP_0$?

a. 0 *c.* n
b. 1 *d.* It cannot be determined.

15. What is the number of permutations of 6 different objects taken all together?

a. 0 *c.* 36
b. 1 *d.* 720

16. What is 0!?

a. 0 *c.* Undefined
b. 1 *d.* 10

17. What is $_nC_n$?

a. 0 *c.* n
b. 1 *d.* It cannot be determined.

Complete the following statements with the best answer.

18. The set of all possible outcomes of a probability experiment is called the ________.

19. The probability of an event can be any number between and including ________ and ________.

20. If an event cannot occur, its probability is ________.

21. The sum of the probabilities of the events in the sample space is ________.

22. When two events cannot occur at the same time, they are said to be ________.

23. When a card is drawn, find the probability of getting

a. A jack *b.* A 4
c. A card less than 6 (an ace is considered above 6)

24. Selecting a Card When a card is drawn from a deck, find the probability of getting

a. A diamond *b.* A 5 or a heart
c. A 5 and a heart *d.* A king
e. A red card

25. Selecting a Sweater At a men's clothing store, 12 men purchased blue golf sweaters, 8 purchased green sweaters, 4 purchased gray sweaters, and 7 bought black sweaters. If a customer is selected at random, find the probability that he purchased

a. A blue sweater
b. A green or gray sweater
c. A green or black or blue sweater
d. A sweater that was not black

26. Rolling Dice When 2 dice are rolled, find the probability of getting

a. A sum of 6 or 7
b. A sum greater than 8
c. A sum less than 3 or greater than 8
d. A sum that is divisible by 3
e. A sum of 16
f. A sum less than 11

27. Appliance Ownership The probability that a person owns a microwave oven is 0.75, that a person owns a compact disk player is 0.25, and that a person owns both a microwave and a CD player is 0.16. Find the probability that a person owns either a microwave or a CD player, but not both.

28. Starting Salaries Of the physics graduates of a university, 30% received a starting salary of $30,000 or more. If 5 of the graduates are selected at random, find the probability that all had a starting salary of $30,000 or more.

29. Selecting Cards Five cards are drawn from an ordinary deck *without* replacement. Find the probability of getting

a. All red cards
b. All diamonds
c. All aces

30. Scholarships The probability that Samantha will be accepted by the college of her choice and obtain a scholarship is 0.35. If the probability that she is accepted by the college is 0.65, find the probability that she will obtain a scholarship given that she is accepted by the college.

31. New-Car Warranty The probability that a customer will buy a new car and an extended warranty is 0.16. If the probability that a customer will purchase a new car is 0.30, find the probability that the customer will also purchase the extended warranty.

32. Bowling and Club Membership Of the members of the Spring Lake Bowling Lanes, 57% have a lifetime membership and bowl regularly (three or more times a week). If 70% of the club members bowl regularly, find the probability that a randomly selected member is a lifetime member, given that he or she bowls regularly.

33. Work and Weather The probability that Mike has to work overtime and it rains is 0.028. Mike hears the weather forecast, and there is a 50% chance of rain. Find the probability that he will have to work overtime, given that it rains.

34. Education of Factory Employees At a large factory, the employees were surveyed and classified according to their level of education and whether they attend a sports event at least once a month. The data are shown in the table.

	Educational level		
Sports event	**High school graduate**	**Two-year college degree**	**Four-year college degree**
Attend	16	20	24
Do not attend	12	19	25

If an employee is selected at random, find the probability that

a. The employee attends sports events regularly, given that he or she graduated from college (2- or 4-year degree)

b. Given that the employee is a high school graduate, he or she does not attend sports events regularly

35. Heart Attacks In a certain high-risk group, the chances of a person having suffered a heart attack are 55%. If 6 people are chosen, find the probability that at least 1 will have had a heart attack.

36. Rolling a Die A single die is rolled 4 times. Find the probability of getting at least one 5.

37. Eye Color If 85% of all people have brown eyes and 6 people are selected at random, find the probability that at least 1 of them has brown eyes.

38. Singer Selection How many ways can 5 sopranos and 4 altos be selected from 7 sopranos and 9 altos?

39. Speaker Selection How many different ways can 8 speakers be seated on a stage?

40. Stocking Machines A soda machine servicer must restock and collect money from 15 machines, each one at a different location. How many ways can she select 4 machines to service in 1 day?

41. ID Cards One company's ID cards consist of 5 letters followed by 2 digits. How many cards can be made if repetitions are allowed? If repetitions are not allowed?

42. Word Permutation How many different arrangements of the letters in the word *number* can be made?

43. Physics Test A physics test consists of 25 true/false questions. How many different possible answer keys can be made?

44. Cellular Telephones How many different ways can 5 cellular telephones be selected from 8 cellular phones?

45. Fruit Selection On a lunch counter, there are 3 oranges, 5 apples, and 2 bananas. If 3 pieces of fruit are selected, find the probability that 1 orange, 1 apple, and 1 banana are selected.

46. Cruise Ship Activities A cruise director schedules 4 different movies, 2 bridge games, and 3 tennis games for a two-day period. If a couple selects 3 activities, find the probability that they attend 2 movies and 1 tennis game.

47. Committee Selection At a sorority meeting, there are 6 seniors, 4 juniors, and 2 sophomores. If a committee of 3 is to be formed, find the probability that 1 of each will be selected.

48. Banquet Meal Choices For a banquet, a committee can select beef, pork, chicken, or veal; baked potatoes or mashed potatoes; and peas or green beans for a vegetable. Draw a tree diagram for all possible choices of a meat, a potato, and a vegetable.

49. Toy Display A toy store manager wants to display 7 identical stuffed dogs, 4 identical stuffed cats, and 3 identical stuffed teddy bears on a shelf. How many different arrangements can be made?

50. Commercial Order A local television station must show commercial X twice, commercial Y twice, and commercial Z three times during a 2-hour show. How many different ways can this be done?

Critical Thinking Challenges

1. **Con Man Game** Consider this problem: A con man has 3 coins. One coin has been specially made and has a head on each side. A second coin has been specially made, and on each side it has a tail. Finally, a third coin has a head and a tail on it. All coins are of the same denomination. The con man places the 3 coins in his pocket, selects one, and shows you one side. It is heads. He is willing to bet you even money that it is the two-headed coin. His reasoning is that it can't be the two-tailed coin since a head is showing; therefore, there is a 50-50 chance of it being the two-headed coin. Would you take the bet?
2. **de Méré Dice Game** Chevalier de Méré won money when he bet unsuspecting patrons that in 4 rolls of 1 die, he could get at least one 6; but he lost money when he bet that in 24 rolls of 2 dice, he could get at least a double 6. Using the probability rules, find the probability of each event and explain why he won the majority of the time on the first game but lost the majority of the time when playing the second game. (*Hint:* Find the probabilities of losing each game and subtract from 1.)
3. **Classical Birthday Problem** How many people do you think need to be in a room so that 2 people will have the same birthday (month and day)? You might think it is 366. This would, of course, guarantee it (excluding leap year), but how many people would need to be in a room so that there would be a 90% probability that 2 people would be born on the same day? What about a 50% probability?

Actually, the number is much smaller than you may think. For example, if you have 50 people in a room, the probability that 2 people will have the same birthday is 97%. If you have 23 people in a room, there is a 50% probability that 2 people were born on the same day!

The problem can be solved by using the probability rules. It must be assumed that all birthdays are equally likely, but this assumption will have little effect on the answers. The way to find the answer is by using the complementary event rule as P(2 people having the same birthday) = 1 − P(all have different birthdays).

For example, suppose there were 3 people in the room. The probability that each had a different birthday would be

$$\frac{365}{365} \cdot \frac{364}{365} \cdot \frac{363}{365} = \frac{_{365}P_3}{365^3} = 0.992$$

Hence, the probability that at least 2 of the 3 people will have the same birthday will be

$$1 - 0.992 = 0.008$$

Hence, for k people, the formula is

$$P(\text{at least 2 people have the same birthday}) = 1 - \frac{_{365}P_k}{365^k}$$

Using your calculator, complete the table and verify that for at least a 50% chance of 2 people having the same birthday, 23 or more people will be needed.

Number of people	Probability that at least 2 have the same birthday
1	0.000
2	0.003
5	0.027
10	
15	
20	
21	
22	
23	

4. **Contracting a Disease** We know that if the probability of an event happening is 100%, then the event is a certainty. Can it be concluded that if there is a 50% chance of contracting a communicable disease through contact with an infected person, there would be a 100% chance of contracting the disease if 2 contacts were made with the infected person? Explain your answer.

Data Projects

1. **Business and Finance** Select a pizza restaurant and a sandwich shop. For the pizza restaurant look at the menu to determine how many sizes, crust types, and toppings are available. How many different pizza types are possible? For the sandwich shop determine how many breads, meats, veggies, cheeses, sauces, and condiments are available. How many different sandwich choices are possible?
2. **Sports and Leisure** When poker games are shown on television, there are often percentages displayed that show how likely it is that a certain hand will win. Investigate how these percentages are determined. Show an example with two competing hands in a Texas Hold 'Em game. Include the percentages that each hand will win after the deal, the flop, the turn, and the river.
3. **Technology** A music player or music organization program can keep track of how many different artists are in a library. First note how many different artists are in your music library. Then find the probability that if 25 songs are selected at random, none will have the same artist.
4. **Health and Wellness** Assume that the gender distribution of babies is such that one-half the time females are born and one-half the time males are born. In a family of 3 children, what is the probability that all are girls? In a family of 4? Is it unusual that in a family with 4 children all would be girls? In a family of 5?
5. **Politics and Economics** Consider the U.S. Senate. Find out about the composition of any three of the Senate's standing committees. How many different committees of Senators are possible, knowing the party composition of the Senate and the number of committee members from each party for each committee?
6. **Your Class** Research the famous Monty Hall probability problem. Conduct a simulation of the Monty Hall problem online using a simulation program or in class using live "contestants." After 50 simulations compare your results to those stated in the research you did. Did your simulation support the conclusions?

Answers to Applying the Concepts

Section 4–1 Tossing a Coin

1. The sample space is the listing of all possible outcomes of the coin toss.
2. The possible outcomes are heads or tails.
3. Classical probability says that a fair coin has a 50% chance of coming up heads and a 50% chance of coming up tails.
4. The law of large numbers says that as you increase the number of trials, the overall results will approach the

theoretical probability. However, since the coin has no "memory," it still has a 50% chance of coming up heads and a 50% chance of coming up tails on the next toss. Knowing what has already happened should not change your opinion on what will happen on the next toss.

5. The empirical approach to probability is based on running an experiment and looking at the results. You cannot do that at this time.
6. Subjective probabilities could be used if you believe the coin is biased.
7. Answers will vary; however, they should address that a fair coin has a 50% chance of coming up heads and a 50% chance of coming up tails on the next flip.

Section 4–2 Which Pain Reliever Is Best?

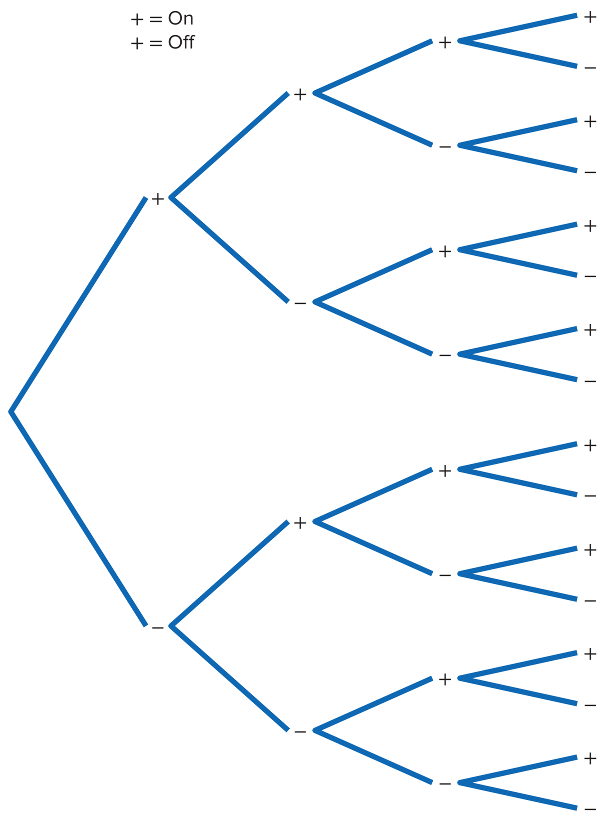
1. There were 192 + 186 + 188 = 566 subjects in the study.
2. The study lasted for 12 weeks.
3. The variables are the type of pain reliever and the side effects.
4. Both variables are qualitative and nominal.
5. The numbers in the table are exact figures.
6. The probability that a randomly selected person was receiving a placebo is 192/566 = 0.339 (about 34%).
7. The probability that a randomly selected person was receiving a placebo or drug A is (192 + 186)/566 = 378/566 = 0.668 (about 67%). These are mutually exclusive events. The complement is that a randomly selected person was receiving drug B.
8. The probability that a randomly selected person was receiving a placebo or experienced a neurological headache is (192 + 55 + 72)/566 = 319/566 = 0.564 (about 56%).
9. The probability that a randomly selected person was not receiving a placebo or experienced a sinus headache is (186 + 188)/566 + 11/566 = 385/566 = 0.680 (about 68%).

Section 4–3 Guilty or Innocent?

1. The probability of another couple with the same characteristics being in that area is $\frac{1}{12}\cdot\frac{1}{10}\cdot\frac{1}{4}\cdot\frac{1}{11}\cdot\frac{1}{3}\cdot\frac{1}{13}\cdot\frac{1}{100}=\frac{1}{20,592,000}$, assuming the characteristics are independent of one another.
2. You would use the multiplication rule, since you are looking for the probability of multiple events happening together.
3. We do not know if the characteristics are dependent or independent, but we assumed independence for the calculation in question 1.
4. The probabilities would change if there were dependence among two or more events.
5. Answers will vary. One possible answer is that probabilities can be used to explain how unlikely it is to have a set of events occur at the same time (in this case, how unlikely it is to have another couple with the same characteristics in that area).
6. Answers will vary. One possible answer is that if the only eyewitness was the woman who was mugged and the probabilities are accurate, it seems very unlikely that a couple matching these characteristics would be in that area at that time. This might cause you to convict the couple.
7. Answers will vary. One possible answer is that our probabilities are theoretical and serve a purpose when appropriate, but that court cases are based on much more than impersonal chance.
8. Answers will vary. One possible answer is that juries decide whether to convict a defendant if they find evidence "beyond a reasonable doubt" that the person is guilty. In probability terms, this means that if the defendant was actually innocent, then the chance of seeing the events that occurred is so unlikely as to have occurred by chance. Therefore, the jury concludes that the defendant is guilty.

Section 4–4 Garage Door Openers

1. Four on/off switches lead to 16 different settings.



2. With 5 on/off switches, there are $2^5 = 32$ different settings. With 6 on/off switches, there are $2^6 = 64$ different settings. In general, if there are k on/off switches, there are 2^k different settings.

3. With 8 consecutive on/off switches, there are $2^8 = 256$ different settings.
4. It is less likely for someone to be able to open your garage door if you have 8 on/off settings (probability about 0.4%) than if you have 4 on/off switches (probability about 6.0%). Having 8 on/off switches in the opener seems pretty safe.
5. Each key blank could be made into $5^5 = 3125$ possible keys.
6. If there were 420,000 Dodge Caravans sold in the United States, then any one key could start about $420{,}000/3125 = 134.4$, or about 134, different Caravans.
7. Answers will vary.

Section 4–5 Counting Rules and Probability

1. There are five different events: each multiple-choice question is an event.
2. These events are independent.
3. If you guess on 1 question, the probability of getting it correct is 0.20. Thus, if you guess on all 5 questions, the probability of getting all of them correct is $(0.20)^5 = 0.00032$.
4. The probability that a person would guess answer A for a question is 0.20, so the probability that a person would guess answer A for each question is $(0.20)^5 = 0.00032$.
5. There are five different events: each matching question is an event.
6. These are dependent events.
7. The probability of getting them all correct if you are guessing is $\frac{1}{5} \cdot \frac{1}{4} \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{1} = \frac{1}{120} = 0.008$.
8. The difference between the two problems is that we are sampling without replacement in the second problem, so the denominator changes in the event probabilities.

5

Discrete Probability Distributions

© Fotosearch RF

STATISTICS TODAY

Is Pooling Worthwhile?

Blood samples are used to screen people for certain diseases. When the disease is rare, health care workers sometimes combine or pool the blood samples of a group of individuals into one batch and then test it. If the test result of the batch is negative, no further testing is needed since none of the individuals in the group has the disease. However, if the test result of the batch is positive, each individual in the group must be tested.

Consider this hypothetical example: Suppose the probability of a person having the disease is 0.05, and a pooled sample of 15 individuals is tested. What is the probability that no further testing will be needed for the individuals in the sample? The answer to this question can be found by using what is called the *binomial distribution*. See Statistics Today—Revisited at the end of the chapter.

This chapter explains probability distributions in general and a specific, often used distribution called the binomial distribution. The Poisson, hypergeometric, geometric, and multinomial distributions are also explained.

OUTLINE

OBJECTIVES

After completing this chapter, you should be able to

1. Construct a probability distribution for a random variable.
2. Find the mean, variance, standard deviation, and expected value for a discrete random variable.
3. Find the exact probability for *X* successes in *n* trials of a binomial experiment.
4. Find the mean, variance, and standard deviation for the variable of a binomial distribution.
5. Find probabilities for outcomes of variables, using the Poisson, hypergeometric, geometric, and multinomial distributions.

Introduction

Many decisions in business, insurance, and other real-life situations are made by assigning probabilities to all possible outcomes pertaining to the situation and then evaluating the results. For example, a saleswoman can compute the probability that she will make 0, 1, 2, or 3 or more sales in a single day. An insurance company might be able to assign probabilities to the number of vehicles a family owns. A self-employed speaker might be able to compute the probabilities for giving 0, 1, 2, 3, or 4 or more speeches each week. Once these probabilities are assigned, statistics such as the mean, variance, and standard deviation can be computed for these events. With these statistics, various decisions can be made. The saleswoman will be able to compute the average number of sales she makes per week, and if she is working on commission, she will be able to approximate her weekly income over a period of time, say, monthly. The public speaker will be able to plan ahead and approximate his average income and expenses. The insurance company can use its information to design special computer forms and programs to accommodate its customers' future needs.

This chapter explains the concepts and applications of what is called a *probability distribution.* In addition, special probability distributions, such as the *binomial, multinomial, Poisson, hypergeometric,* and *geometric* distributions, are explained.

5–1 Probability Distributions

OBJECTIVE 1

Construct a probability distribution for a random variable.

Before probability distribution is defined formally, the definition of a variable is reviewed. In Chapter 1, a *variable* was defined as a characteristic or attribute that can assume different values. Various letters of the alphabet, such as *X, Y,* or *Z,* are used to represent variables. Since the variables in this chapter are associated with probability, they are called *random variables.*

For example, if a die is rolled, a letter such as *X* can be used to represent the outcomes. Then the value that *X* can assume is 1, 2, 3, 4, 5, or 6, corresponding to the outcomes of rolling a single die. If two coins are tossed, a letter, say *Y,* can be used to represent the number of heads, in this case 0, 1, or 2. As another example, if the temperature at 8:00 A.M. is 43° and at noon it is 53°, then the values *T* that the temperature assumes are said to be random, since they are due to various atmospheric conditions at the time the temperature was taken.

A **random variable** is a variable whose values are determined by chance.

Also recall from Chapter 1 that you can classify variables as discrete or continuous by observing the values the variable can assume. If a variable can assume only a specific number of values, such as the outcomes for the roll of a die or the outcomes for the toss of a coin, then the variable is called a *discrete variable.*

Discrete variables have a finite number of possible values or an infinite number of values that can be counted. The word *counted* means that they can be enumerated using the numbers 1, 2, 3, etc. For example, the number of joggers in Riverview Park each day and the number of phone calls received after a TV commercial airs are examples of discrete variables, since they can be counted.

Variables that can assume all values in the interval between any two given values are called *continuous variables.* For example, if the temperature goes from 62° to 78° in a 24-hour period, it has passed through every possible number from 62 to 78. *Continuous random variables are obtained from data that can be measured rather than counted.* Continuous random variables can assume an infinite number of values and can be decimal and fractional values. On a continuous scale, a person's weight might be exactly 183.426 pounds if a scale could measure weight to the thousandths place; however, on a

digital scale that measures only to tenths of a pound, the weight would be 183.4 pounds. Examples of continuous variables are heights, weights, temperatures, and time. In this chapter only discrete random variables are used; Chapter 6 explains continuous random variables.

The procedure shown here for constructing a probability distribution for a discrete random variable uses the probability experiment of tossing three coins. Recall that when three coins are tossed, the sample space is represented as TTT, TTH, THT, HTT, HHT, HTH, THH, HHH; and if X is the random variable for the number of heads, then X assumes the value 0, 1, 2, or 3.

Probabilities for the values of X can be determined as follows:

No heads	One head			Two heads			Three heads
TTT	TTH	THT	HTT	HHT	HTH	THH	HHH
$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$
$\frac{1}{8}$	$\frac{3}{8}$			$\frac{3}{8}$			$\frac{1}{8}$

Hence, the probability of getting no heads is $\frac{1}{8}$, one head is $\frac{3}{8}$, two heads is $\frac{3}{8}$, and three heads is $\frac{1}{8}$. From these values, a probability distribution can be constructed by listing the outcomes and assigning the probability of each outcome, as shown here.

Number of heads *X*	0	1	2	3
Probability *P(X)*	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

A **discrete probability distribution** consists of the values a random variable can assume and the corresponding probabilities of the values. The probabilities are determined theoretically or by observation.

Procedure Table

Constructing a Probability Distribution

Step 1 Make a frequency distribution for the outcomes of the variable.

Step 2 Find the probability for each outcome by dividing the frequency of the outcome by the sum of the frequencies.

Step 3 If a graph is required, place the outcomes on the x axis and the probabilities on the y axis, and draw vertical bars for each outcome and its corresponding probability.

Discrete probability distributions can be shown by using a graph or a table. Probability distributions can also be represented by a formula. See Exercises 31–36 at the end of this section for examples.

EXAMPLE 5–1 Rolling a Die

Construct a probability distribution for rolling a single die.

SOLUTION

Since the sample space is 1, 2, 3, 4, 5, 6 and each outcome has a probability of $\frac{1}{6}$, the distribution is as shown.

Outcome *X*	1	2	3	4	5	6
Probability *P(X)*	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

When probability distributions are shown graphically, the values of X are placed on the x axis and the probabilities $P(X)$ on the y axis. These graphs are helpful in determining the shape of the distribution (right-skewed, left-skewed, or symmetric).

EXAMPLE 5–2 Tossing Coins

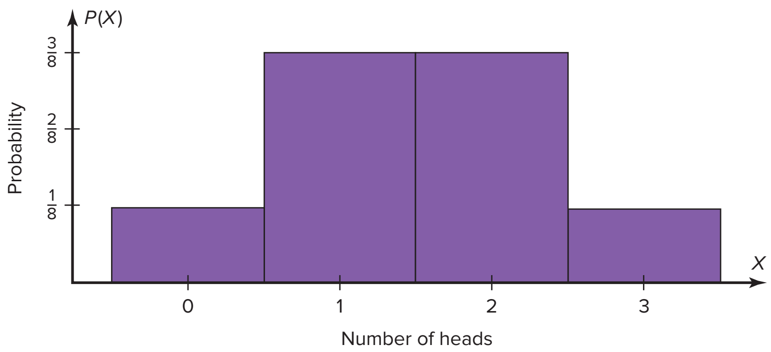
Represent graphically the probability distribution for the sample space for tossing three coins.

Number of heads X	0	1	2	3
Probability $P(X)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

SOLUTION

The values that X assumes are located on the x axis, and the values for $P(X)$ are located on the y axis. The graph is shown in Figure 5–1.

FIGURE 5–1 Probability Distribution for Example 5–2



Note that for visual appearances, it is not necessary to start with 0 at the origin.

Examples 5–1 and 5–2 are illustrations of *theoretical* probability distributions. You did not need to actually perform the experiments to compute the probabilities. In contrast, to construct actual probability distributions, you must observe the variable over a period of time. They are empirical, as shown in Example 5–3.

EXAMPLE 5–3 Battery Packages

A convenience store sells AA batteries in 2 per package, 4 per package, 6 per package, and 8 per package. The store sells 5 two-packs, 10 four-packs, 8 six-packs, and 2 eight-packs over the weekend. Construct a probability distribution and draw a graph for the variable.

SOLUTION

Step 1 Make a frequency distribution for the variable.

Outcome	2	4	6	8
Frequency	5	10	8	2

Step 2 Find the probability for each outcome. The total of the frequencies is 25. Hence,

$$P(2) = \frac{5}{25} = 0.20 \qquad P(4) = \frac{10}{25} = 0.40$$

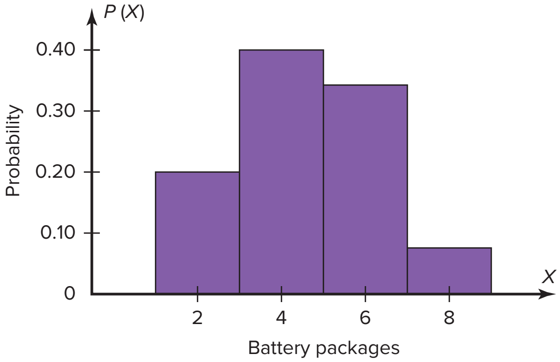
$$P(6) = \frac{8}{25} = 0.32 \qquad P(8) = \frac{2}{25} = 0.08$$

The probability distribution is

Outcome *X*	2	4	6	8
Probability *P*(*X*)	0.20	0.40	0.32	0.08

Step 3 Draw the graph, using the x axis for the outcomes and the y axis for the probabilities. See Figure 5–2.

FIGURE 5–2
Probability Distribution for Example 5–3



Two Requirements for a Probability Distribution

1. The sum of the probabilities of all the events in the sample space must equal 1; that is, $\Sigma P(X) = 1$.
2. The probability of each event in the sample space must be between or equal to 0 and 1. That is, $0 \leq P(X) \leq 1$.

The first requirement states that the sum of the probabilities of all the events must be equal to 1. This sum cannot be less than 1 or greater than 1 since the sample space includes *all* possible outcomes of the probability experiment. The second requirement states that the probability of any individual event must be a value from 0 to 1. The reason (as stated in Chapter 4) is that the range of the probability of any individual value can be 0, 1, or any value between 0 and 1. A probability cannot be a negative number or greater than 1.

EXAMPLE 5–4 Probability Distributions

Determine whether each distribution is a probability distribution.

a.

X	2	4	6	8	10
P(*X*)	0.3	0.4	0.1	0.2	0.1

b.

X	♠	♦	♥	♣
P(*X*)	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{8}$	$\frac{1}{8}$

c.

X	BB	BG	GB	GG
P(*X*)	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

d.

X	3	7	10	12
P(*X*)	−0.6	0.3	0.3	0.2

SPEAKING OF STATISTICS Coins, Births, and Other Random (?) Events

Examples of random events such as tossing coins are used in almost all books on probability. But is flipping a coin really a random event?

Tossing coins dates back to ancient Roman times when the coins usually consisted of the Emperor's head on one side (i.e., heads) and another icon such as a ship on the other side (i.e., tails). Tossing coins was used in both fortune telling and ancient Roman games.

A Chinese form of divination called the *I-Ching* (pronounced E-Ching) is thought to be at least 4000 years old. It consists of 64 hexagrams made up of six horizontal lines. Each line is either broken or unbroken, representing the yin and the yang. These 64 hexagrams are supposed to represent all possible situations in life. To consult the *I-Ching,* a question is asked and then three coins are tossed six times. The way the coins fall, either heads up or heads down, determines whether the line is broken (yin) or unbroken (yang). Once the hexagram is determined, its meaning is consulted and interpreted to get the answer to the question. (*Note:* Another method used to determine the hexagram employs yarrow sticks.)

© PNC/Getty Images RF

In the 16th century, a mathematician named Abraham DeMoivre used the outcomes of tossing coins to study what later became known as the normal distribution; however, his work at that time was not widely known.

Mathematicians usually consider the outcomes of a coin toss to be a random event. That is, each probability of getting a head is $\frac{1}{2}$, and the probability of getting a tail is $\frac{1}{2}$. Also, it is not possible to predict with 100% certainty which outcome will occur. But new studies question this theory. During World War II a South African mathematician named John Kerrich tossed a coin 10,000 times while he was interned in a German prison camp. Although the results of his experiment were never officially recorded, most references indicate that out of his 10,000 tosses, 5,067 were heads.

Several studies have shown that when a coin-tossing device is used, the probability that a coin will land on the same side on which it is placed on the coin-tossing device is about 51%. It would take about 10,000 tosses to become aware of this bias. Furthermore, researchers showed that when a coin is spun on its edge, the coin falls tails up about 80% of the time since there is more metal on the heads side of a coin. This makes the coin slightly heavier on the heads side than on the tails side.

Another assumption commonly made in probability theory is that the number of male births is equal to the number of female births and that the probability of a boy being born is $\frac{1}{2}$ and the probability of a girl being born is $\frac{1}{2}$. We know this is not exactly true.

In the later 1700s, a French mathematician named Pierre Simon Laplace attempted to prove that more males than females are born. He used records from 1745 to 1770 in Paris and showed that the percentage of females born was about 49%. Although these percentages vary somewhat from location to location, further surveys show they are generally true worldwide. Even though there are discrepancies, we generally consider the outcomes to be 50-50 since these discrepancies are relatively small.

Based on this article, would you consider the coin toss at the beginning of a football game fair?

SOLUTION

a. No. The sum of the probabilities is greater than 1.

b. Yes. The sum of the probabilities of all the events is equal to 1. Each probability is greater than or equal to 0 and less than or equal to 1.

c. Yes. The sum of the probabilities of all the events is equal to 1. Each probability is greater than or equal to 0 and less than or equal to 1.

d. No. One of the probabilities is less than 0.

Many variables in business, education, engineering, and other areas can be analyzed by using probability distributions. Section 5–2 shows methods for finding the mean and standard deviation for a probability distribution.

Applying the Concepts 5–1

Dropping College Courses

Use the following table to answer the questions.

Reason for dropping a college course	Frequency	Percentage
Too difficult	45	
Illness	40	
Change in work schedule	20	
Change of major	14	
Family-related problems	9	
Money	7	
Miscellaneous	6	
No meaningful reason	3	

1. What is the variable under study? Is it a random variable?
2. How many people were in the study?
3. Complete the table.
4. From the information given, what is the probability that a student will drop a class because of illness? Money? Change of major?
5. Would you consider the information in the table to be a probability distribution?
6. Are the categories mutually exclusive?
7. Are the categories independent?
8. Are the categories exhaustive?
9. Are the two requirements for a discrete probability distribution met?

See page 309 for the answers.

Exercises 5–1

1. Define and give three examples of a random variable.

2. Explain the difference between a discrete and a continuous random variable.

3. Give three examples of a discrete random variable.

4. Give three examples of a continuous random variable.

5. List three continuous random variables and three discrete random variables associated with a major league baseball game.

6. What is a probability distribution? Give an example.

For Exercises 7 through 12, determine whether the distribution represents a probability distribution. If it does not, state why.

7.

X	15	16	20	25
***P*(*X*)**	0.2	0.5	0.7	−0.8

8.

X	5	7	9
***P*(*X*)**	0.6	0.8	−0.4

9.

X	−5	−3	0	2	4
***P*(*X*)**	0.1	0.3	0.2	0.3	0.1

10.

X	20	30	40	50
***P*(*X*)**	0.05	0.35	0.4	0.2

11.

X	3	6	9	1
***P*(*X*)**	0.3	0.4	0.3	0.1

12.

X	3	7	9	12	14
$P(X)$	$\frac{4}{13}$	$\frac{1}{13}$	$\frac{3}{13}$	$\frac{1}{13}$	$\frac{2}{13}$

For Exercises 13 through 18, state whether the variable is discrete or continuous.

13. The number of books in your school's library

14. The number of people who play the state lottery each day

15. The temperature of the water in Lake Erie

16. The time it takes to have a medical physical exam

17. The total number of points scored in a basketball game

18. The blood pressures of all patients admitted to a hospital on a specific day

For Exercises 19 through 26, construct a probability distribution for the data and draw a graph for the distribution.

19. Statistical Calculators The probability that a college bookstore sells 0, 1, 2, or 3 statistical calculators on any given day is $\frac{4}{9}$, $\frac{2}{9}$, $\frac{2}{9}$, and $\frac{1}{9}$, respectively.

20. Investment Return The probabilities of a return on an investment of \$5000, \$7000, and \$9000 are $\frac{1}{2}$, $\frac{3}{8}$ and $\frac{1}{8}$, respectively.

21. Automobile Tires The probability that an automobile repair shop sells 0, 1, 2, 3, or 4 tires on any given day is 0.25, 0.05, 0.30, 0.00, and 0.40 respectively.

22. DVD Rentals The probabilities that a customer will rent 0, 1, 2, 3, or 4 DVDs on a single visit to the rental store are 0.15, 0.25, 0.3, 0.25, and 0.05, respectively.

23. Loaded Die A die is loaded in such a way that the probabilities of getting 1, 2, 3, 4, 5, and 6 are $\frac{1}{2}$, $\frac{1}{6}$, $\frac{1}{12}$, $\frac{1}{12}$, $\frac{1}{12}$, and $\frac{1}{12}$, respectively.

24. Item Selection The probabilities that a customer selects 1, 2, 3, 4, and 5 items at a convenience store are 0.32, 0.12, 0.23, 0.18, and 0.15, respectively.

25. Student Classes The probabilities that a student is registered for 2, 3, 4, or 5 classes are 0.01, 0.34, 0.62, and 0.03, respectively.

26. Garage Space The probabilities that a randomly selected home has garage space for 0, 1, 2, or 3 cars are 0.22, 0.33, 0.37, and 0.08, respectively.

27. Triangular Numbers The first six triangular numbers (1, 3, 6, 10, 15, 21) are printed one each on one side of a card. The cards are placed face down and mixed. Choose two cards at random, and let X be the sum of the two numbers. Construct the probability distribution for this random variable X.

28. Child Play in Day Care In a popular day care center, the probability that a child will play with the computer is 0.45; the probability that he or she will play dress-up is 0.27; play with blocks, 0.18; and paint, 0.1. Construct the probability distribution for this discrete random variable.

29. Goals in Hockey The probability that a hockey team scores a total of 1 goal in a game is 0.124; 2 goals, 0.297; 3 goals, 0.402; 4 goals, 0.094; and 5 goals, 0.083. Construct the probability distribution for this discrete random variable and draw the graph.

30. Mathematics Tutoring Center At a drop-in mathematics tutoring center, each teacher sees 4 to 8 students per hour. The probability that a tutor sees 4 students in an hour is 0.117; 5 students, 0.123; 6 students, 0.295; and 7 students, 0.328. Find the probability that a tutor sees 8 students in an hour, construct the probability distribution, and draw the graph.

Extending the Concepts

A probability distribution can be written in formula notation such as $P(X) = 1/X$, where $X = 2, 3, 6$. The distribution is shown as follows:

X	2	3	6
$P(X)$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$

For Exercises 31 through 36, write the distribution for the formula and determine whether it is a probability distribution.

31. $P(X) = X/6$ for $X = 1, 2, 3$

32. $P(X) = X$ for $X = 0.2, 0.3, 0.5$

33. $P(X) = X/6$ for $X = 3, 4, 7$

34. $P(X) = X + 0.1$ for $X = 0.1, 0.02, 0.04$

35. $P(X) = X/7$ for $X = 1, 2, 4$

36. $P(X) = X/(X + 2)$ for $X = 0, 1, 2$

37. Computer Games The probability that a child plays one computer game is one-half as likely as that of playing two computer games. The probability of playing three games is twice as likely as that of playing two games, and the probability of playing four games is the average of the other three. Let X be the number of computer games played. Construct the probability distribution for this random variable and draw the graph.

5–2 Mean, Variance, Standard Deviation, and Expectation

OBJECTIVE 2

Find the mean, variance, standard deviation, and expected value for a discrete random variable.

The mean, variance, and standard deviation for a probability distribution are computed differently from the mean, variance, and standard deviation for samples. This section explains how these measures—as well as a new measure called the *expectation*—are calculated for probability distributions.

Mean

In Chapter 3, the mean for a sample or population was computed by adding the values and dividing by the total number of values, as shown in these formulas:

$$\text{Sample mean:} \quad \overline{X} = \frac{\Sigma X}{n} \qquad \text{Population mean:} \quad \mu = \frac{\Sigma X}{N}$$

But how would you compute the mean of the number of spots that show on top when a die is rolled? You could try rolling the die, say, 10 times, recording the number of spots, and finding the mean; however, this answer would only approximate the true mean. What about 50 rolls or 100 rolls? Actually, the more times the die is rolled, the better the approximation. You might ask, then, How many times must the die be rolled to get the exact answer? *It must be rolled an infinite number of times.* Since this task is impossible, the previous formulas cannot be used because the denominators would be infinity. Hence, a new method of computing the mean is necessary. This method gives the exact theoretical value of the mean as if it were possible to roll the die an infinite number of times.

Historical Note

A professor, Augustin Louis Cauchy (1789–1857), wrote a book on probability. While he was teaching at the Military School of Paris, one of his students was Napoleon Bonaparte.

Before the formula is stated, an example will be used to explain the concept. Suppose two coins are tossed repeatedly, and the number of heads that occurred is recorded. What will be the mean of the number of heads? The sample space is

HH, HT, TH, TT

and each outcome has a probability of $\frac{1}{4}$. Now, in the long run, you would *expect* two heads (HH) to occur approximately $\frac{1}{4}$ of the time, one head to occur approximately $\frac{1}{2}$ of the time (HT or TH), and no heads (TT) to occur approximately $\frac{1}{4}$ of the time. Hence, on average, you would expect the number of heads to be

$$2 \cdot \tfrac{1}{4} + 1 \cdot \tfrac{1}{2} + 0 \cdot \tfrac{1}{4} = 1$$

That is, if it were possible to toss the coins many times or an infinite number of times, the *average* of the number of heads would be 1.

Hence, to find the mean for a probability distribution, you must multiply each possible outcome by its corresponding probability and find the sum of the products.

Formula for the Mean of a Probability Distribution

The mean of a random variable with a discrete probability distribution is

$$\begin{aligned}\mu &= X_1 \cdot P(X_1) + X_2 \cdot P(X_2) + X_3 \cdot P(X_3) + \cdots + X_n \cdot P(X_n)\\ &= \Sigma X \cdot P(X)\end{aligned}$$

where $X_1, X_2, X_3, \ldots, X_n$ are the outcomes and $P(X_1), P(X_2), P(X_3), \ldots, P(X_n)$ are the corresponding probabilities.

Note: $\Sigma X \cdot P(X)$ means to sum the products.

Rounding Rule for the Mean, Variance, and Standard Deviation for a Probability Distribution The rounding rule for the mean, variance, and standard deviation for variables of a probability distribution is this: The mean, variance, and standard deviation should be rounded to one more decimal place than the outcome X. When fractions are used, they should be reduced to lowest terms.

Examples 5–5 through 5–8 illustrate the use of the formula.

EXAMPLE 5–5 Rolling a Die

Find the mean of the number of spots that appear when a die is tossed.

SOLUTION

In the toss of a die, the mean can be computed thus.

Outcome *X*	1	2	3	4	5	6
Probability *P*(*X*)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$\mu = \Sigma X \cdot P(X) = 1 \cdot \tfrac{1}{6} + 2 \cdot \tfrac{1}{6} + 3 \cdot \tfrac{1}{6} + 4 \cdot \tfrac{1}{6} + 5 \cdot \tfrac{1}{6} + 6 \cdot \tfrac{1}{6}$$
$$= \tfrac{21}{6} = 3\tfrac{1}{2} \text{ or } 3.5$$

That is, when a die is tossed many times, the theoretical mean will be 3.5. Note that even though the die cannot show a 3.5, the theoretical average is 3.5.

The reason why this formula gives the theoretical mean is that in the long run, each outcome would occur approximately $\frac{1}{6}$ of the time. Hence, multiplying the outcome by its corresponding probability and finding the sum would yield the theoretical mean. In other words, outcome 1 would occur approximately $\frac{1}{6}$ of the time, outcome 2 would occur approximately $\frac{1}{6}$ of the time, etc.

EXAMPLE 5–6 Children in a Family

In families with four children, find the mean number of children who will be girls.

SOLUTION

First, it is necessary to find the sample space. There are 16 outcomes, as shown.

BBBB	BBGG	GGGG
BBBG	BGBG	GGGB
BBGB	GGBB	GGBG
BGBB	GBGB	GBGG
GBBB	BGGB	BGGG
GBBG		

(A tree diagram may help.)

Next, make a probability distribution.

Number of girls *X*	0	1	2	3	4
Probability *P*(*X*)	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

Then multiply X and $P(X)$ for each outcome and find the sum.

$$\mu = \Sigma X \cdot P(X) = 0 \cdot \frac{1}{16} + 1 \cdot \frac{4}{16} + 2 \cdot \frac{6}{16} + 3 \cdot \frac{4}{16} + 4 \cdot \frac{1}{16} = 2$$

Hence, the mean of the number of females is 2.

EXAMPLE 5–7 Tossing Coins

If three coins are tossed, find the mean of the number of heads that occur. (See the table preceding Example 5–1.)

SOLUTION

The probability distribution is

Number of heads *X*	0	1	2	3
Probability *P*(*X*)	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

The mean is

$$\mu = \Sigma X \cdot P(X) = 0 \cdot \tfrac{1}{8} + 1 \cdot \tfrac{3}{8} + 2 \cdot \tfrac{3}{8} + 3 \cdot \tfrac{1}{8} = \tfrac{12}{8} = 1\tfrac{1}{2} \text{ or } 1.5$$

The value 1.5 cannot occur as an outcome. Nevertheless, it is the long-run or theoretical average.

EXAMPLE 5–8 Battery Packages

Find the mean of the number of batteries sold over the weekend at a convenience store. See Example 5–3.

SOLUTION

The probability distribution is

Outcome *X*	2	4	6	8
Probability *P*(*X*)	0.20	0.40	0.32	0.08

$$\mu = \Sigma X \cdot P(X) = 2(0.20) + 4(0.40) + 6(0.32) + 8(0.08) = 4.56$$

Hence, the mean number of batteries sold is 4.56.

Historical Note

Fey Manufacturing Co., located in San Francisco, invented the first three-reel, automatic payout slot machine in 1895.

Variance and Standard Deviation

For a probability distribution, the mean of the random variable describes the measure of the so-called long-run or theoretical average, but it does not tell anything about the spread of the distribution. Recall from Chapter 3 that to measure this spread or variability, statisticians use the variance and standard deviation. These formulas were used:

$$\sigma^2 = \frac{\Sigma(X-\mu)^2}{N} \qquad \text{or} \qquad \sigma = \sqrt{\frac{\Sigma(X-\mu)^2}{N}}$$

These formulas cannot be used for a random variable of a probability distribution since N is infinite, so the variance and standard deviation must be computed differently.

To find the variance for the random variable of a probability distribution, subtract the theoretical mean of the random variable from each outcome and square the difference. Then multiply each difference by its corresponding probability and add the products. The formula is

$$\sigma^2 = \Sigma[(X-\mu)^2 \cdot P(X)]$$

Finding the variance by using this formula is somewhat tedious. So for simplified computations, a shortcut formula can be used. This formula is algebraically equivalent to the longer one and is used in the examples that follow.

Formula for the Variance of a Probability Distribution

Find the variance of a probability distribution by multiplying the square of each outcome by its corresponding probability, summing those products, and subtracting the square of the mean. The formula for the variance of a probability distribution is

$$\sigma^2 = \Sigma[X^2 \cdot P(X)] - \mu^2$$

The standard deviation of a probability distribution is

$$\sigma = \sqrt{\sigma^2} \qquad \text{or} \qquad \sigma = \sqrt{\Sigma[X^2 \cdot P(X)] - \mu^2}$$

Remember that the variance and standard deviation cannot be negative.

EXAMPLE 5–9 Rolling a Die

Compute the variance and standard deviation for the probability distribution in Example 5–5.

SOLUTION

Recall that the mean is $\mu = 3.5$, as computed in Example 5–5. Square each outcome and multiply by the corresponding probability, sum those products, and then subtract the square of the mean.

$$\sigma^2 = (1^2 \cdot \tfrac{1}{6} + 2^2 \cdot \tfrac{1}{6} + 3^2 \cdot \tfrac{1}{6} + 4^2 \cdot \tfrac{1}{6} + 5^2 \cdot \tfrac{1}{6} + 6^2 \cdot \tfrac{1}{6}) - (3.5)^2 = 2.917$$

To get the standard deviation, find the square root of the variance.

$$\sigma = \sqrt{2.917} \approx 1.708$$

Hence, the standard deviation for rolling a die is 1.708.

EXAMPLE 5–10 Selecting Numbered Balls

A box contains 5 balls. Two are numbered 3, one is numbered 4, and two are numbered 5. The balls are mixed and one is selected at random. After a ball is selected, its number is recorded. Then it is replaced. If the experiment is repeated many times, find the variance and standard deviation of the numbers on the balls.

SOLUTION

Let X be the number on each ball. The probability distribution is

Number of ball *X*	3	4	5
Probability *P*(*X*)	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{2}{5}$

The mean is

$$\mu = \Sigma X \cdot P(X) = 3 \cdot \tfrac{2}{5} + 4 \cdot \tfrac{1}{5} + 5 \cdot \tfrac{2}{5} = 4$$

The variance is

$$\begin{aligned}\sigma &= \Sigma[X^2 \cdot P(X)] - \mu^2 \\ &= 3^2 \cdot \tfrac{2}{5} + 4^2 \cdot \tfrac{1}{5} + 5^2 \cdot \tfrac{2}{5} - 4^2 \\ &= 16\tfrac{4}{5} - 16 \\ &= \tfrac{4}{5} \text{ or } 0.8\end{aligned}$$

The standard deviation is

$$\sigma = \sqrt{\frac{4}{5}} = \sqrt{0.8} \approx 0.894$$

The mean, variance, and standard deviation can also be found by using vertical columns, as shown.

X	*P*(*X*)	*X* · *P*(*X*)	$X^2 \cdot P(X)$
3	0.4	1.2	3.6
4	0.2	0.8	3.2
5	0.4	2.0	10
		$\Sigma X \cdot P(X) = 4.0$	16.8

Find the mean by summing the $X \cdot P(X)$ column, and find the variance by summing the $X^2 \cdot P(X)$ column and subtracting the square of the mean.

$$\sigma^2 = 16.8 - 4^2 = 16.8 - 16 = 0.8$$

and

$$\sigma = \sqrt{0.8} \approx 0.894$$

EXAMPLE 5–11 On Hold for Talk Radio

A talk radio station has four telephone lines. If the host is unable to talk (i.e., during a commercial) or is talking to a person, the other callers are placed on hold. When all lines are in use, others who are trying to call in get a busy signal. The probability that 0, 1, 2, 3, or 4 people will get through is shown in the probability distribution. Find the variance and standard deviation for the distribution.

X	0	1	2	3	4
P(X)	0.18	0.34	0.23	0.21	0.04

Should the station have considered getting more phone lines installed?

SOLUTION

The mean is

$$\begin{aligned}\mu &= \Sigma X \cdot P(X)\\ &= 0 \cdot (0.18) + 1 \cdot (0.34) + 2 \cdot (0.23) + 3 \cdot (0.21) + 4 \cdot (0.04)\\ &= 1.59\end{aligned}$$

The variance is

$$\begin{aligned}\sigma^2 &= \Sigma[X^2 \cdot P(X)] - \mu^2\\ &= [0^2 \cdot (0.18) + 1^2 \cdot (0.34) + 2^2 \cdot (0.23) + 3^2 \cdot (0.21) + 4^2 \cdot (0.04)] - 1.59^2\\ &= (0 + 0.34 + 0.92 + 1.89 + 0.64) - 2.528\\ &= 3.79 - 2.528 = 1.262\end{aligned}$$

The standard deviation is $\sigma = \sqrt{\sigma^2}$, or $\sigma = \sqrt{1.262} \approx 1.123$.

No. The mean number of people calling at any one time is 1.59. Since the standard deviation is 1.123, most callers would be accommodated by having four phone lines because $\mu + 2\sigma$ would be $1.59 + 2(1.123) = 3.836 \approx 4.0$. Very few callers would get a busy signal since at least 75% of the callers would either get through or be put on hold. (See Chebyshev's theorem in Section 3–2.)

Expectation

Another concept related to the mean for a probability distribution is that of expected value or expectation. Expected value is used in various types of games of chance, in insurance, and in other areas, such as decision theory.

The **expected value** of a discrete random variable of a probability distribution is the theoretical average of the variable. The formula is

$$\mu = E(X) = \Sigma X \cdot P(X)$$

The symbol $E(X)$ is used for the expected value.

The formula for the expected value is the same as the formula for the theoretical mean. The expected value, then, is the theoretical mean of the probability distribution. That is, $E(X) = \mu$.

When expected value problems involve money, it is customary to round the answer to the nearest cent.

EXAMPLE 5–12 Winning Tickets

One thousand tickets are sold at $1 each for a smart television valued at $750. What is the expected value of the gain if you purchase one ticket?

SOLUTION

The problem can be set up as follows:

	Win	Lose
Gain *X*	\$749	−\$1
Probability *P(X)*	$\frac{1}{1000}$	$\frac{999}{1000}$

Two things should be noted. First, for a win, the net gain is $749, since you do not get the cost of the ticket ($1) back. Second, for a loss, the gain is represented by a negative number, in this case −$1. The solution, then, is

$$E(X) = \$749 \cdot \frac{1}{1000} + (-\$1) \cdot \frac{999}{1000} = -\$0.25$$

Hence, a person would lose, on average, −$0.25 on each ticket purchased.

Expected value problems of this type can also be solved by finding the overall gain (i.e., the value of the prize won or the amount of money won, not considering the cost of the ticket for the prize or the cost to play the game) and subtracting the cost of the tickets or the cost to play the game, as shown:

$$E(X) = \$750 \cdot \frac{1}{1000} - \$1 = -\$0.25$$

Here, the overall gain ($750) must be used.

Note that the expectation is −$0.25. This does not mean that you lose $0.25, since you can only win a television set valued at $750 or lose $1 on the ticket. What this expectation means is that the average of the losses is $0.25 for each of the 1000 ticket holders. Here is another way of looking at this situation: If you purchased one ticket each week over a long time, the average loss would be $0.25 per ticket, since theoretically, on average, you would win the television set once for each 1000 tickets purchased.

EXAMPLE 5–13 UNO Cards

Ten cards are selected from a deck of UNO cards. There are 2 cards numbered 0; 1 card numbered 2; 3 cards numbered 4; 2 cards numbered 8; and 2 cards numbered 9. If the cards are mixed up and one card is selected at random, find the expected value of the card.

SOLUTION

The probability for the standard distribution for the cards is

Value *X*	0	2	4	8	9
Probability *P*(*X*)	$\frac{2}{10}$	$\frac{1}{10}$	$\frac{3}{10}$	$\frac{2}{10}$	$\frac{2}{10}$

The expected value is

$$\mu = \Sigma X \cdot P(X) = 0 \cdot \frac{2}{10} + 2 \cdot \frac{1}{10} + 4 \cdot \frac{3}{10} + 8 \cdot \frac{2}{10} + 9 \cdot \frac{2}{10} = 4\frac{4}{5} = 4.8$$

EXAMPLE 5–14 Bond Investment

A financial adviser suggests that his client select one of two types of bonds in which to invest \$5000. Bond X pays a return of 4% and has a default rate of 2%. Bond Y has a $2\frac{1}{2}$% return and a default rate of 1%. Find the expected rate of return and decide which bond would be a better investment. When the bond defaults, the investor loses all the investment.

SOLUTION

The return on bond X is \$5000 · 4% = \$200. The expected return then is

$$E(X) = \$200(0.98) - \$5000(0.02) = \$96$$

The return on bond Y is \$5000 · $2\frac{1}{2}$% = \$125. The expected return then is

$$E(Y) = \$125(0.99) - \$5000(0.01) = \$73.75$$

Hence, bond X would be a better investment since the expected return is higher.

In gambling games, if the expected value of the game is zero, the game is said to be fair. If the expected value of a game is positive, then the game is in favor of the player. That is, the player has a better than even chance of winning. If the expected value of the game is negative, then the game is said to be in favor of the house. That is, in the long run, the players will lose money.

In his book *Probabilities in Everyday Life* (Ivy Books, 1986), author John D. McGervy gives the expectations for various casino games. For keno, the house wins \$0.27 on every \$1.00 bet. For Chuck-a-Luck, the house wins about \$0.52 on every \$1.00 bet. For roulette, the house wins about \$0.90 on every \$1.00 bet. For craps, the house wins about \$0.88 on every \$1.00 bet. The bottom line here is that if you gamble long enough, sooner or later you will end up losing money.

Applying the Concepts 5–2

Radiation Exposure

On March 28, 1979, the nuclear generating facility at Three Mile Island, Pennsylvania, began discharging radiation into the atmosphere. People exposed to even low levels of radiation can experience health problems ranging from very mild to severe, even causing death. A local newspaper reported that 11 babies were born with kidney problems in the three-county area surrounding the Three Mile Island nuclear power plant. The expected value for that problem in infants in that area was 3. Answer the following questions.

1. What does *expected value* mean?
2. Would you expect the exact value of 3 all the time?

3. If a news reporter stated that the number of cases of kidney problems in newborns was nearly four times as many as was usually expected, do you think pregnant mothers living in that area would be overly concerned?
4. Is it unlikely that 11 occurred by chance?
5. Are there any other statistics that could better inform the public?
6. Assume that 3 out of 2500 babies were born with kidney problems in that three-county area the year before the accident. Also assume that 11 out of 2500 babies were born with kidney problems in that three-county area the year after the accident. What is the real percentage increase in that abnormality?
7. Do you think that pregnant mothers living in that area should be overly concerned after looking at the results in terms of rates?

See page 309 for the answers.

Exercises 5–2

1. **Coffee with Meals** A researcher wishes to determine the number of cups of coffee a customer drinks with an evening meal at a restaurant. Find the mean, variance, and standard deviation for the distribution.

X	0	1	2	3	4
$P(X)$	0.31	0.42	0.21	0.04	0.02

2. **Suit Sales** The number of suits sold per day at a retail store is shown in the table, with the corresponding probabilities. Find the mean, variance, and standard deviation of the distribution.

Number of suits sold X	19	20	21	22	23
Probability $P(X)$	0.2	0.2	0.3	0.2	0.1

If the manager of the retail store wants to be sure that he has enough suits for the next 5 days, how many should the manager purchase?

3. **Daily Newspapers** A survey was taken of the number of daily newspapers a person reads per day. Find the mean, variance, and standard deviation of the distribution.

X	0	1	2	3
$P(X)$	0.42	0.35	0.20	0.03

4. **Trivia Quiz** The probabilities that a player will get 5 to 10 questions right on a trivia quiz are shown below. Find the mean, variance, and standard deviation for the distribution.

X	5	6	7	8	9	10
$P(X)$	0.05	0.2	0.4	0.1	0.15	0.1

5. **New Homes** A contractor has four new home plans. Plan 1 is a home with six windows. Plan 2 is a home with seven windows. Plan 3 has eight windows, and plan 4 has nine windows. The probability distribution for the sale of the homes is shown. Find the mean, variance, and standard deviation for the number of windows in the homes that the contractor builds.

X	6	7	8	9
$P(X)$	0.3	0.4	0.25	0.05

6. **Traffic Accidents** The county highway department recorded the following probabilities for the number of accidents per day on a certain freeway for one month. The number of accidents per day and their corresponding probabilities are shown. Find the mean, variance, and standard deviation.

Number of accidents X	0	1	2	3	4
Probability $P(X)$	0.4	0.2	0.2	0.1	0.1

7. **Fitness Machine** A fitness center bought a new exercise machine called the Mountain Climber. They decided to keep track of how many people used the machine over a 3-hour period. Find the mean, variance, and standard deviation for the probability distribution. Here X is the number of people who used the machine.

X	0	1	2	3	4
$P(X)$	0.1	0.2	0.4	0.2	0.1

8. **Benford's Law** The leading digits in actual data, such as stock prices, population numbers, death rates, and lengths of rivers, do not occur randomly as one might suppose, but instead follow a distribution according to Benford's law. Below is the probability distribution for the leading digits in real-life lists of data. Calculate the mean for the distribution.

X	1	2	3	4	5	6	7	8	9
$P(X)$	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046

9. **Automobiles** A survey shows the probability of the number of automobiles that families in a certain housing

plan own. Find the mean, variance, and standard deviation for the distribution.

X	1	2	3	4	5
$P(X)$	0.27	0.46	0.21	0.05	0.01

10. **Pizza Deliveries** A pizza shop owner determines the number of pizzas that are delivered each day. Find the mean, variance, and standard deviation for the distribution shown. If the manager stated that 45 pizzas were delivered on one day, do you think that this is a believable claim?

Number of deliveries X	35	36	37	38	39
Probability $P(X)$	0.1	0.2	0.3	0.3	0.1

11. **Grab Bags** A convenience store has made up 20 grab bag gifts and is offering them for \$2.00 a bag. Ten bags contain merchandise worth \$1.00. Six bags contain merchandise worth \$2.00, and four bags contain merchandise worth \$3.00. Suppose you purchase one bag. What is your expected gain or loss?

12. **Job Bids** A landscape contractor bids on jobs where he can make \$3000 profit. The probabilities of getting 1, 2, 3, or 4 jobs per month are shown.

Number of jobs	1	2	3	4
Probability	0.2	0.3	0.4	0.1

Find the contractor's expected profit per month.

13. **Rolling Dice** If a person rolls doubles when she tosses two dice, she wins \$5. For the game to be fair, how much should she pay to play the game?

14. **Dice Game** A person pays \$2 to play a certain game by rolling a single die once. If a 1 or a 2 comes up, the person wins nothing. If, however, the player rolls a 3, 4, 5, or 6, he or she wins the difference between the number rolled and \$2. Find the expectation for this game. Is the game fair?

15. **Lottery Prizes** A lottery offers one \$1000 prize, one \$500 prize, and five \$100 prizes. One thousand tickets are sold at \$3 each. Find the expectation if a person buys one ticket.

16. **Lottery Prizes** In Exercise 15, find the expectation if a person buys two tickets. Assume that the player's ticket is replaced after each draw and that the same ticket can win more than one prize.

17. **Winning the Lottery** For a daily lottery, a person selects a three-digit number. If the person plays for \$1, she can win \$500. Find the expectation. In the same daily lottery, if a person boxes a number, she will win \$80. Find the expectation if the number 123 is played for \$1 and boxed. (When a number is "boxed," it can win when the digits occur in any order.)

18. **Life Insurance** A 35-year-old woman purchases a \$100,000 term life insurance policy for an annual payment of \$360. Based on a period life table for the U.S. government, the probability that she will survive the year is 0.999057. Find the expected value of the policy for the insurance company.

19. **Roulette** A roulette wheel has 38 numbers, 1 through 36, 0, and 00. One-half of the numbers from 1 through 36 are red, and the other half are black; 0 and 00 are green. A ball is rolled, and it falls into one of the 38 slots, giving a number and a color. The payoffs (winnings) for a \$1 bet are as follows:

Red or black	\$1	0	\$35
Odd or even	\$1	00	\$35
1–18	\$1	Any single number	\$35
9–36	\$1	0 or 00	\$17

If a person bets \$1, find the expected value for each.

a. Red
b. Even
c. 00
d. Any single number
e. 0 or 00

Extending the Concepts

20. **Rolling Dice** Construct a probability distribution for the sum shown on the faces when two dice are rolled. Find the mean, variance, and standard deviation of the distribution.

21. **Rolling a Die** When one die is rolled, the expected value of the number of dots is 3.5. In Exercise 20, the mean number of dots was found for rolling two dice. What is the mean number of dots if three dice are rolled?

22. The formula for finding the variance for a probability distribution is

$$\sigma^2 = \Sigma[(X - \mu)^2 \cdot P(X)]$$

Verify algebraically that this formula gives the same result as the shortcut formula shown in this section.

23. Complete the following probability distribution if $P(6)$ equals two-thirds of $P(4)$. Then find μ, σ^2, and σ for the distribution.

X	1	2	4	6	9
$P(X)$	0.23	0.18	?	?	0.015

24. **Rolling Two Dice** Roll two dice 100 times and find the mean, variance, and standard deviation of the sum of the dots. Compare the result with the theoretical results obtained in Exercise 20.

25. **Extracurricular Activities** Conduct a survey of the number of extracurricular activities your classmates are enrolled in. Construct a probability distribution and find the mean, variance, and standard deviation.

26. Promotional Campaign In a recent promotional campaign, a company offered these prizes and the corresponding probabilities. Find the expected value of winning. The tickets are free.

Number of prizes	Amount	Probability
1	\$100,000	$\frac{1}{1,000,000}$
2	10,000	$\frac{1}{50,000}$
5	1,000	$\frac{1}{10,000}$
10	100	$\frac{1}{1000}$

If the winner has to mail in the winning ticket to claim the prize, what will be the expectation if the cost of the stamp is considered? Use the current cost of a stamp for a first-class letter.

27. Probability Distribution A bag contains five balls numbered 1, 2, 4, 7, and *. Choose two balls at random without replacement and add the numbers. If one ball has the *, double the amount on the other ball. Construct the probability distribution for this random variable X and calculate μ, σ^2, and σ.

Technology Step by Step

TI-84 Plus Step by Step

Calculating the Mean and Variance of a Discrete Random Variable

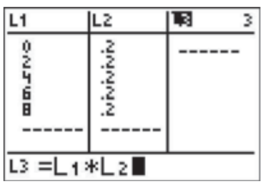
To calculate the mean and variance for a discrete random variable by using the formulas:

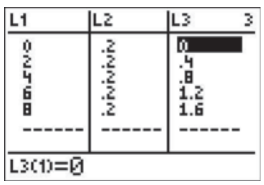
1. Enter the x values into L_1 and the probabilities into L_2.
2. Move the cursor to the top of the L_3 column so that L_3 is highlighted.
3. Type $\mathbf{L_1}$ multiplied by $\mathbf{L_2}$, then press **ENTER**.
4. Move the cursor to the top of the L_4 column so that L_4 is highlighted.
5. Type $\mathbf{L_1}$ followed by the x^2 key multiplied by $\mathbf{L_2}$, then press **ENTER**.
6. Type **2nd QUIT** to return to the home screen.
7. Type **2nd LIST**, move the cursor to MATH, type 5 for sum, then type $\mathbf{L_3}$, then press **ENTER**. (This is the mean.)
8. Type **2nd ENTER**, move the cursor to L_3, type $\mathbf{L_4}$, then press **ENTER**.

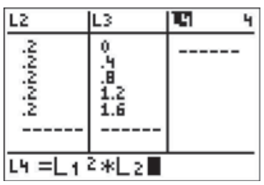
Example TI5–1

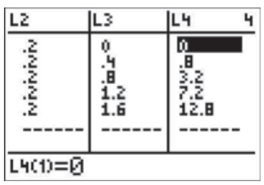
Number on ball X	0	2	4	6	8
Probability $P(X)$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$

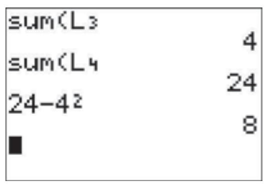
Using the data from Example TI5–1 gives the following:





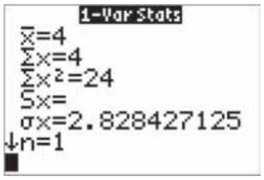






The mean is 4 and the variance is 8.

To calculate the mean and standard deviation for a discrete random variable without using the formulas, modify the procedure to calculate the mean and standard deviation from grouped data (Chapter 3) by entering the x values into L_1 and the probabilities into L_2.



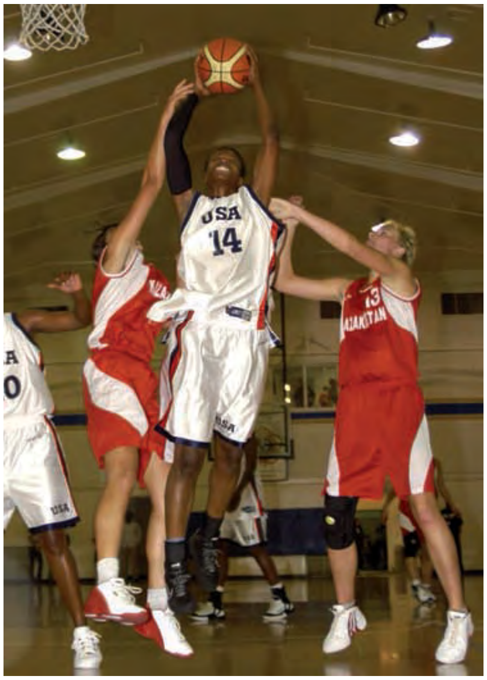
The mean is 4 and the standard deviation is 2.828427125. To calculate the variance, square the standard deviation.

5–3 The Binomial Distribution

OBJECTIVE 3

Find the exact probability for X successes in n trials of a binomial experiment.

Many types of probability problems have only two outcomes or can be reduced to two outcomes. For example, when a coin is tossed, it can land heads or tails. When a 6-sided die is rolled, it will either land on an odd or even number. In a basketball game, a team either wins or loses. A true/false item can be answered in only two ways, true or false. Other situations can be reduced to two outcomes. For example, a medical treatment can be classified as effective or ineffective, depending on the results. A person can be classified as having normal or abnormal blood pressure, depending on the measure of the blood pressure gauge. A multiple-choice question, even though there are four or five answer choices, can be classified as correct or incorrect. Situations like these are called *binomial experiments.*



U.S. Navy photo

Each repetition of the experiment is called a *trial*.

Historical Note

In 1653, Blaise Pascal created a triangle of numbers called *Pascal's triangle* that can be used in the binomial distribution.

A **binomial experiment** is a probability experiment that satisfies the following four requirements:

1. There must be a fixed number of trials.
2. Each trial can have only two outcomes or outcomes that can be reduced to two outcomes. These outcomes can be considered as either success or failure.
3. The outcomes of each trial must be independent of one another.
4. The probability of a success must remain the same for each trial.

The word *success* does not imply that something good or positive has occurred. For example, in a probability experiment, we might want to select 10 people and let S represent the number of people who were in an automobile accident in the last six months. In this case, a success would not be a positive or good thing.

EXAMPLE 5–15

Decide whether each experiment is a binomial experiment. If not, state the reason why.

a. Selecting 20 university students and recording their class rank

b. Selecting 20 students from a university and recording their gender

c. Drawing five cards from a deck without replacement and recording whether they are red or black cards

d. Selecting five students from a large school and asking them if they are on the dean's list

e. Recording the number of children in 50 randomly selected families

SOLUTION

a. No. There are five possible outcomes: freshman, sophomore, junior, senior, and graduate student.

b. Yes. All four requirements are met.

c. No. Since the cards are not replaced, the events are not independent.

d. Yes. All four requirements are met.

e. No. There can be more than two categories for the answers.

A binomial experiment and its results give rise to a special probability distribution called the *binomial distribution.*

The outcomes of a binomial experiment and the corresponding probabilities of these outcomes are called a **binomial distribution.**

In binomial experiments, the outcomes are usually classified as successes or failures. For example, the correct answer to a multiple-choice item can be classified as a success, but any of the other choices would be incorrect and hence classified as a failure. The notation that is commonly used for binomial experiments and the binomial distribution is defined now.

Notation for the Binomial Distribution

$P(S)$	The symbol for the probability of success
$P(F)$	The symbol for the probability of failure
p	The numerical probability of a success
q	The numerical probability of a failure

$$P(S) = p \qquad \text{and} \qquad P(F) = 1 - p = q$$

n	The number of trials
X	The number of successes in n trials

Note that $0 \le X \le n$ and $X = 0, 1, 2, 3, \ldots, n$.

The probability of a success in a binomial experiment can be computed with this formula.

Binomial Probability Formula

In a binomial experiment, the probability of exactly X successes in n trials is

$$P(X) = \frac{n!}{(n-X)!X!} \cdot p^X \cdot q^{n-X}$$

An explanation of why the formula works is given following Example 5–16.

EXAMPLE 5–16 Tossing Coins

A coin is tossed 3 times. Find the probability of getting exactly two heads.

SOLUTION

This problem can be solved by looking at the sample space. There are three ways to get two heads.

HHH, <u>HHT, HTH, THH,</u> TTH, THT, HTT, TTT

The answer is $\frac{3}{8}$, or 0.375.

Looking at the problem in Example 5–16 from the standpoint of a binomial experiment, one can show that it meets the four requirements.

1. There are a fixed number of trials (three).
2. There are only two outcomes for each trial, heads or tails.
3. The outcomes are independent of one another (the outcome of one toss in no way affects the outcome of another toss).
4. The probability of a success (heads) is $\frac{1}{2}$ in each case.

In this case, $n = 3$, $X = 2$, $p = \frac{1}{2}$, and $q = \frac{1}{2}$. Hence, substituting in the formula gives

$$P(2\text{ heads}) = \frac{3!}{(3-2)!2!} \cdot \left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^1 = \frac{3}{8} = 0.375$$

which is the same answer obtained by using the sample space.

The same example can be used to explain the formula. First, note that there are three ways to get exactly two heads and one tail from a possible eight ways. They are HHT, HTH, and THH. In this case, then, the number of ways of obtaining two heads from three coin tosses is ${}_3C_2$, or 3, as shown in Chapter 4. In general, the number of ways to get X successes from n trials without regard to order is

$${}_nC_X = \frac{n!}{(n-X)!X!}$$

This is the first part of the binomial formula. (Some calculators can be used for this.)

Next, each success has a probability of $\frac{1}{2}$ and can occur twice. Likewise, each failure has a probability of $\frac{1}{2}$ and can occur once, giving the $(\frac{1}{2})^2(\frac{1}{2})^1$ part of the formula. To generalize, then, each success has a probability of p and can occur X times, and each failure has a probability of q and can occur $n - X$ times. Putting it all together yields the binomial probability formula.

When sampling is done without replacement, such as in surveys, the events are dependent events; however, the events can be considered independent if the size of the sample is no more than 5% of the size of the population. That is, $n \le 0.05N$. The reason is that when one item is selected from a large number of items and is not replaced before the second item is selected, the change in the probability of the second item being selected is so small that it can be ignored.

EXAMPLE 5–17 Survey on Doctor Visits

A survey found that one out of five Americans says he or she has visited a doctor in any given month. If 10 people are selected at random, find the probability that exactly 3 will have visited a doctor last month.

Source: Reader's Digest.

SOLUTION

In this case, $n = 10$, $X = 3$, $p = \frac{1}{5}$, and $q = \frac{4}{5}$. Hence,

$$P(3) = \frac{10!}{(10-3)!3!}\left(\frac{1}{5}\right)^3\left(\frac{4}{5}\right)^7 \approx 0.201$$

So, there is a 0.201 probability that in a random sample of 10 people, exactly 3 of them visited a doctor in the last month.

EXAMPLE 5–18 Survey on Employment

A survey from Teenage Research Unlimited (Northbrook, Illinois) found that 30% of teenage consumers receive their spending money from part-time jobs. If 5 teenagers are selected at random, find the probability that at least 3 of them will have part-time jobs.

SOLUTION

To find the probability that at least 3 have part-time jobs, it is necessary to find the individual probabilities for 3, or 4, or 5 and then add them to get the total probability.

$$P(3) = \frac{5!}{(5-3)!3!}(0.3)^3(0.7)^2 \approx 0.132$$

$$P(4) = \frac{5!}{(5-4)!4!}(0.3)^4(0.7)^1 \approx 0.028$$

$$P(5) = \frac{5!}{(5-5)!5!}(0.3)^5(0.7)^0 \approx 0.002$$

Hence,

$$\begin{aligned}&P(\text{at least three teenagers have part-time jobs})\\&\quad = 0.132 + 0.028 + 0.002 = 0.162\end{aligned}$$

Computing probabilities by using the binomial probability formula can be quite tedious at times, so tables have been developed for selected values of n and p. Table B in Appendix A gives the probabilities for individual events. Example 5–19 shows how to use Table B to compute probabilities for binomial experiments.

EXAMPLE 5–19 Tossing Coins

Solve the problem in Example 5–16 by using Table B.

SOLUTION

Since $n = 3$, $X = 2$, and $p = 0.5$, the value 0.375 is found as shown in Figure 5–3.

FIGURE 5–3 Using Table B for Example 5–16

								p						
n	X	0.05	0.1	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.75	0.8	0.9	0.95
2	0													
	1													
	2													
3	0							0.125						
	1							0.375						
	2							0.375						
	3							0.125						

$p = 0.5$

$n = 3$

$X = 2$

EXAMPLE 5–20 Survey on Fear of Being Home Alone at Night

Public Opinion reported that 5% of Americans are afraid of being alone in a house at night. If a random sample of 20 Americans is selected, find these probabilities by using the binomial table.

a. There are exactly 5 people in the sample who are afraid of being alone at night.

b. There are at most 3 people in the sample who are afraid of being alone at night.

c. There are at least 3 people in the sample who are afraid of being alone at night.

Source: 100% American by Daniel Evan Weiss.

SOLUTION

a. $n = 20$, $p = 0.05$, and $X = 5$. From the table, we get 0.002.

b. $n = 20$ and $p = 0.05$. "At most 3 people" means 0, or 1, or 2, or 3.

Hence, the solution is

$$P(X \le 3) = P(0) + P(1) + P(2) + P(3) = 0.358 + 0.377 + 0.189 + 0.060$$
$$= 0.984$$

c. $n = 20$ and $p = 0.05$. "At least 3 people" means 3, 4, 5, . . . , 20. This problem can best be solved by finding $P(X \le 2) = P(0) + P(1) + P(2)$ and subtracting from 1.

$$P(0) + P(1) + P(2) = 0.358 + 0.377 + 0.189 = 0.924$$
$$P(X \ge 3) = 1 - 0.924 = 0.076$$

EXAMPLE 5–21 Driving While Intoxicated

A report from the Secretary of Health and Human Services stated that 70% of single-vehicle traffic fatalities that occur at night on weekends involve an intoxicated driver. If a sample of 15 single-vehicle traffic fatalities that occur on a weekend night is selected, find the probability that between 10 and 15, inclusive, accidents involved drivers who were intoxicated.

SOLUTION

$n = 15$, $p = 0.70$, and X is 10, 11, 12, 13, 14, or 15.

For $X = 10$,	$P(10) = 0.206$	For $X = 13$,	$P(13) = 0.092$
For $X = 11$,	$P(11) = 0.219$	For $X = 14$,	$P(14) = 0.031$
For $X = 12$,	$P(12) = 0.170$	For $X = 15$,	$P(15) = 0.005$

$$P(10 \leq X \leq 15) = 0.206 + 0.219 + 0.170 + 0.092 + 0.031 + 0.005 = 0.723$$

Hence, the probability that between 10 and 15 drivers, inclusive, were intoxicated is 0.723.

The mean, variance, and standard deviation for a binomial variable can be found by using the formulas in Section 5–2; however, shorter but mathematically equivalent formulas for the mean, variance, and standard deviation are used.

OBJECTIVE 4

Find the mean, variance, and standard deviation for the variable of a binomial distribution.

Mean, Variance, and Standard Deviation for the Binomial Distribution

The mean, variance, and standard deviation of a variable that has the *binomial distribution* can be found by using the following formulas.

Mean: $\mu = n \cdot p$ Variance: $\sigma^2 = n \cdot p \cdot q$ Standard deviation: $\sigma = \sqrt{n \cdot p \cdot q}$

EXAMPLE 5–22 Tossing a Coin

A coin is tossed 4 times. Find the mean, variance, and standard deviation of the number of heads that will be obtained.

SOLUTION

With the formulas for the binomial distribution and $n = 4$, $p = \frac{1}{2}$, and $q = \frac{1}{2}$, the results are

$$\mu = n \cdot p = 4 \cdot \tfrac{1}{2} = 2$$

$$\sigma^2 = n \cdot p \cdot q = 4 \cdot \tfrac{1}{2} \cdot \tfrac{1}{2} = 1$$

$$\sigma = \sqrt{1} = 1$$

In this case, the mean is two heads. The variance is 1 and the standard deviation is 1.

From Example 5–22, when four coins are tossed many, many times, the average of the number of heads that appear is 2, and the standard deviation of the number of heads is 1. Note that these are theoretical values.

As stated previously, this problem can be solved by using the formulas for expected value. The distribution is shown.

No. of heads *X*	0	1	2	3	4
Probability *P(X)*	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

$$\mu = E(X) = \Sigma X \cdot P(X) = 0 \cdot \tfrac{1}{16} + 1 \cdot \tfrac{4}{16} + 2 \cdot \tfrac{6}{16} + 3 \cdot \tfrac{4}{16} + 4 \cdot \tfrac{1}{16} = \tfrac{32}{16} = 2$$

$$\sigma^2 = \Sigma X^2 \cdot P(X) - \mu^2$$

$$= 0^2 \cdot \tfrac{1}{16} + 1^2 \cdot \tfrac{4}{16} + 2^2 \cdot \tfrac{6}{16} + 3^2 \cdot \tfrac{4}{16} + 4^2 \cdot \tfrac{1}{16} - 2^2 = \tfrac{80}{16} - 4 = 1$$

$$\sigma = \sqrt{1} = 1$$

Hence, the simplified binomial formulas give the same results.

EXAMPLE 5–23 Rolling a Die

An 8-sided die (with the numbers 1 through 8 on the faces) is rolled 560 times. Find the mean, variance, and standard deviation of the number of 7s that will be rolled.

SOLUTION

This is a binomial experiment with $n = 560$, $p = \frac{1}{8}$, and $q = \frac{7}{8}$ so that

$$\mu = n \cdot p = 560 \cdot \tfrac{1}{8} = 70$$

$$\sigma^2 = n \cdot p \cdot q = 560 \cdot \tfrac{1}{8} \cdot \tfrac{7}{8} = 61\tfrac{1}{4} = 61.25$$

$$\sigma = \sqrt{61.25} = 7.826$$

In this case, the mean of the number of 7s obtained is 70. The variance is 61.25, and the standard deviation is 7.826.

EXAMPLE 5–24 Intoxicated Drivers

The *Sourcebook of Criminal Justice Statistics* states that 65% of Americans favor sentencing drunk drivers to jail even if they have not caused an accident. If a random number of 1000 individuals is selected, find the mean, variance, and standard deviation of the people who feel this way.

SOLUTION

This is a binomial situation since either people feel that drunk drivers should be sentenced or they feel that they should not.

$$\mu = n \cdot p = (1000)(0.65) = 650$$

$$\sigma^2 = n \cdot p \cdot q = (1000)(0.65)(0.35) = 227.5$$

$$\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{227.5} = 15.083$$

For the sample, the mean is 650, the variance is 277.5, and the standard deviation is 15.083.

Applying the Concepts 5–3

Unsanitary Restaurants

Health officials routinely check the sanitary condition of restaurants. Assume you visit a popular tourist spot and read in the newspaper that in 3 out of every 7 restaurants checked, unsatisfactory health conditions were found. Assuming you are planning to eat out 10 times while you are there on vacation, answer the following questions.

1. How likely is it that you will eat at three restaurants with unsanitary conditions?
2. How likely is it that you will eat at four or five restaurants with unsanitary conditions?
3. Explain how you would compute the probability of eating in at least one restaurant with unsanitary conditions. Could you use the complement to solve this problem?
4. What is the most likely number to occur in this experiment?
5. How variable will the data be around the most likely number?
6. How do you know that this is a binomial distribution?
7. If it is a binomial distribution, does that mean that the likelihood of a success is always 50% since there are only two possible outcomes?

Check your answers by using the following computer-generated table.

Mean = 4.29 Std. dev. = 1.56492

X	$P(X)$	Cum. prob.
0	0.00371	0.00371
1	0.02784	0.03155
2	0.09396	0.12552
3	0.18792	0.31344
4	0.24665	0.56009
5	0.22199	0.78208
6	0.13874	0.92082
7	0.05946	0.98028
8	0.01672	0.99700
9	0.00279	0.99979
10	0.00021	1.00000

See pages 309–310 for the answers.

Exercises 5–3

1. Which of the following are binomial experiments or can be reduced to binomial experiments?

a. Surveying 100 people to determine if they like Sudsy Soap

b. Tossing a coin 100 times to see how many heads occur

c. Drawing a card with replacement from a deck and getting a heart

d. Asking 1000 people which brand of cigarettes they smoke

e. Testing four different brands of aspirin to see which brands are effective

2. Which of the following are binomial experiments or can be reduced to binomial experiments?

a. Testing one brand of aspirin by using 10 people to determine whether it is effective

b. Asking 100 people if they smoke

c. Checking 1000 applicants to see whether they were admitted to White Oak College

d. Surveying 300 prisoners to see how many different crimes they were convicted of

e. Surveying 300 prisoners to see whether this is their first offense

3. Compute the probability of X successes, using Table B in Appendix A.

a. $n = 2, p = 0.30, X = 1$
b. $n = 4, p = 0.60, X = 3$
c. $n = 5, p = 0.10, X = 0$
d. $n = 10, p = 0.40, X = 4$
e. $n = 12, p = 0.90, X = 2$

4. Compute the probability of X successes, using Table B in Appendix A.

a. $n = 15, p = 0.80, X = 12$
b. $n = 17, p = 0.05, X = 0$
c. $n = 20, p = 0.50, X = 10$
d. $n = 16, p = 0.20, X = 3$

5. Compute the probability of X successes, using the binomial formula.

a. $n = 6, X = 3, p = 0.03$
b. $n = 4, X = 2, p = 0.18$
c. $n = 5, X = 3, p = 0.63$

6. Compute the probability of X successes, using the binomial formula.

a. $n = 9, X = 0, p = 0.42$
b. $n = 10, X = 5, p = 0.37$

*For Exercises 7 through 16, assume all variables are binomial. (*Note: *If values are not found in Table B of Appendix A, use the binomial formula.)*

7. Belief in UFOs A survey found that 10% of Americans believe that they have seen a UFO. For a sample of 10 people, find each probability:

a. That at least 2 people believe that they have seen a UFO
b. That 2 or 3 people believe that they have seen a UFO
c. That exactly 1 person believes that he or she has seen a UFO

8. Multiple-Choice Exam A student takes a 20-question, multiple-choice exam with five choices for each question and guesses on each question. Find the probability of guessing at least 15 out of 20 correctly. Would you consider this event likely or unlikely to occur? Explain your answer.

9. High Blood Pressure Twenty percent of Americans ages 25 to 74 have high blood pressure. If 16 randomly selected Americans ages 25 to 74 are selected, find each probability.

a. None will have high blood pressure.
b. One-half will have high blood pressure.
c. Exactly 4 will have high blood pressure.

Source: www.factfinder.census.gov

10. High School Dropouts Approximately 10.3% of American high school students drop out of school before graduation. Choose 10 students entering high school at random. Find the probability that

a. No more than 2 drop out
b. At least 6 graduate
c. All 10 stay in school and graduate

Source: www.infoplease.com.

11. Advertising Three out of four people think most advertising seeks to persuade people to buy things they don't need or can't afford. Find the probability that exactly 5 out of 9 randomly selected people will agree with this statement.

Source: Opinion Research Corporation.

12. Language Spoken at Home by the U.S. Population In 2014 the percentage of the U.S. population that speak English only in the home is 78.9%. Choose 15 U.S. people at random. What is the probability that exactly one-third of them speak English only? At least one-third? What is the probability that at least 9 do not speak English in the home?

Source: World Almanac

13. Prison Inmates Forty percent of prison inmates were unemployed when they entered prison. If 5 inmates are randomly selected, find these probabilities:

a. Exactly 3 were unemployed.
b. At most 4 were unemployed.
c. At least 3 were unemployed.
d. Fewer than 2 were unemployed.

Source: U.S. Department of Justice.

14. Destination Weddings Twenty-six percent of couples who plan to marry this year are planning destination weddings. In a random sample of 12 couples who plan to marry, find the probability that

a. Exactly 6 couples will have a destination wedding
b. At least 6 couples will have a destination wedding
c. Fewer than 5 couples will have a destination wedding

Source: Time magazine.

15. People Who Have Some College Education Fifty-three percent of all persons in the U.S. population have at least some college education. Choose 10 persons at random. Find the probability that

a. Exactly one-half have some college education
b. At least 5 do not have any college education
c. Fewer than 5 have some college education

Source: New York Times Almanac.

16. Guidance Missile System A missile guidance system has 5 fail-safe components. The probability of each failing is 0.05. Find these probabilities.

a. Exactly 2 will fail.
b. More than 2 will fail.

c. All will fail.

d. Compare the answers for parts *a*, *b*, and *c*, and explain why these results are reasonable.

17. Find the mean, variance, and standard deviation for each of the values of *n* and *p* when the conditions for the binomial distribution are met.

a. $n = 100, p = 0.75$

b. $n = 300, p = 0.3$

c. $n = 20, p = 0.5$

d. $n = 10, p = 0.8$

18. Find the mean, variance, and standard deviation for each of the values of *n* and *p* when the conditions for the binomial distributions are met.

a. $n = 1000, p = 0.1$

b. $n = 500, p = 0.25$

c. $n = 50, p = \frac{2}{5}$

d. $n = 36, p = \frac{1}{6}$

19. Airline Accidents Twenty-five percent of commercial airline accidents are caused by bad weather. If 300 commercial accidents are randomly selected, find the mean, variance, and standard deviation of the number of accidents caused by bad weather.

Source: The New York Times.

20. Tossing Coins Find the mean, variance, and standard deviation for the number of heads when 10 coins are tossed.

21. American and Foreign-Born Citizens In 2014 the percentage of the U.S. population who was foreign-born was 13.1. Choose 60 U.S. residents at random. How many would you expect to be American-born? Find the mean, variance, and standard deviation for the number who are foreign-born.

Source: World Almanac 2012.

22. Federal Government Employee E-mail Use It has been reported that 83% of federal government employees use e-mail. If a sample of 200 federal government employees is selected, find the mean, variance, and standard deviation of the number who use e-mail.

Source: USA TODAY.

23. Watching Fireworks A survey found that 21% of Americans watch fireworks on television on July 4. Find the mean, variance, and standard deviation of the number of individuals who watch fireworks on television on July 4 if a random sample of 1000 Americans is selected.

Source: USA Snapshot, *USA TODAY.*

24. Alternate Sources of Fuel Eighty-five percent of Americans favor spending government money to develop alternative sources of fuel for automobiles. For a random sample of 120 Americans, find the mean, variance, and standard deviation for the number who favor government spending for alternative fuels.

Source: www.pollingreport.com

25. Survey on Bathing Pets A survey found that 25% of pet owners had their pets bathed professionally rather than do it themselves. If 18 pet owners are randomly selected, find the probability that exactly 5 people have their pets bathed professionally.

Source: USA Snapshot, *USA TODAY.*

26. Survey on Answering Machine Ownership In a survey, 63% of Americans said they own an answering machine. If 14 Americans are selected at random, find the probability that exactly 9 own an answering machine.

Source: USA Snapshot, *USA TODAY.*

27. Poverty and the Federal Government One out of every three Americans believes that the U.S. government should take "primary responsibility" for eliminating poverty in the United States. If 10 Americans are selected, find the probability that at most 3 will believe that the U.S. government should take primary responsibility for eliminating poverty.

Source: Harper's Index.

28. Internet Purchases Thirty-two percent of adult Internet users have purchased products or services online. For a random sample of 200 adult Internet users, find the mean, variance, and standard deviation for the number who have purchased goods or services online.

Source: www.infoplease.com

29. Runaways Fifty-eight percent of runaways in the United Stated are female. In 20 runaways are selected at random, find the probability that exactly 14 are female.

Source: U.S. Department of Justice.

30. Job Elimination In a recent year, 13% of businesses have eliminated jobs. If 5 businesses are selected at random, find the probability that at least 3 have eliminated jobs during that year.

Source: USA TODAY.

31. Survey of High School Seniors Of graduating high school seniors, 14% said that their generation will be remembered for their social concerns. If 7 graduating seniors are selected at random, find the probability that either 2 or 3 will agree with that statement.

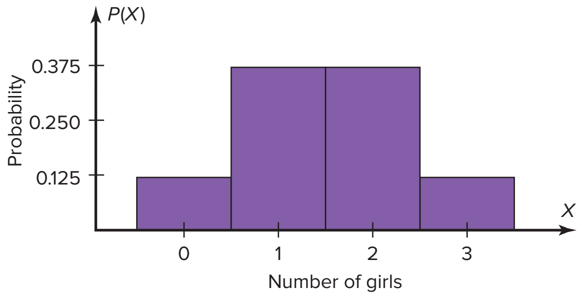
Source: USA TODAY.

32. Is this a binomial distribution? Explain.

X	0	1	2	3
$P(X)$	0.064	0.288	0.432	0.216

Extending the Concepts

33. Children in a Family The graph shown here represents the probability distribution for the number of girls in a family of three children. From this graph, construct a probability distribution.



34. Construct a binomial distribution graph for the number of defective computer chips in a lot of 4 if $p = 0.3$.

35. Show that the mean for a binomial random variable X with $n = 3$ is $3p$.

Technology Step by Step

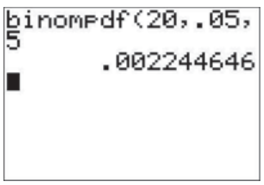
TI-84 Plus Step by Step

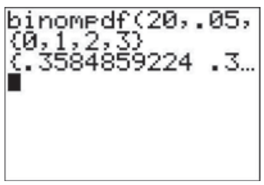
Binomial Random Variables

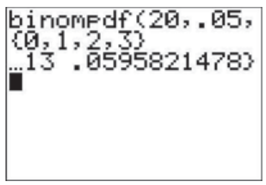
To find the probability for a binomial variable:
Press **2nd [DISTR]** then **A (ALPHA MATH)** for binompdf.
The form is binompdf(n,p,X). On some calculators, you will have a menu showing "trials" (n), "p", and x-value (X). After inputting the values, you will select **PASTE** and press **ENTER**. This will then show binompdf(n,p,X) for the values you entered.

Example: $n = 20$, $X = 5$, $p = .05$ (Example 5–20*a* from the text)
binompdf(20,.05,5), then press **ENTER** for the probability.

Example: $n = 20$, $X = 0, 1, 2, 3$, $p = .05$ (Example 5–20*b* from the text).
binompdf(20,.05,{0,1,2,3}), then press **ENTER.**
The calculator will display the probabilities in a list. Use the arrow keys to view the entire display.

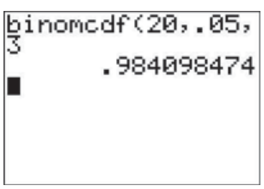






To find the cumulative probability for a binomial random variable:
Press **2nd [DISTR]** then **B (ALPHA APPS)** for binomcdf
The form is binomcdf(n,p,X). This will calculate the cumulative probability for values from 0 to X.

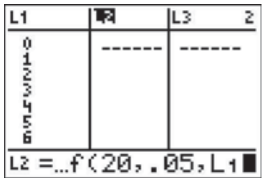
Example: $n = 20$, $X = 0, 1, 2, 3$, $p = .05$ (Example 5–20*b* from the text)
binomcdf(20,.05,3), then press **ENTER.**

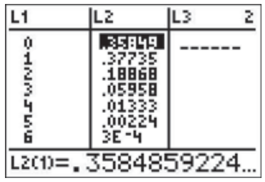


To construct a binomial probability table:

1. Enter the X values (0 through n) into L_1.
2. Move the cursor to the top of the L_2 column so that L_2 is highlighted.
3. Type the command binompdf(n,p,L_1), then press **ENTER.**

Example: $n = 20$, $p = .05$ (Example 5–20 from the text)





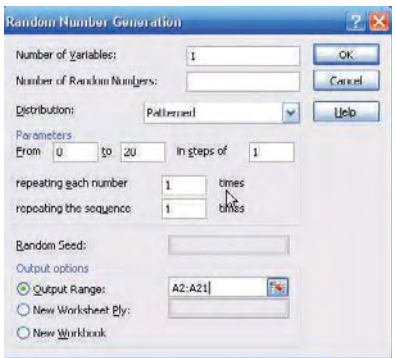
EXCEL
Step by Step

Creating a Binomial Distribution and Graph

These instructions will demonstrate how Excel can be used to construct a binomial distribution table for $n = 20$ and $p = 0.35$.

1. Type **X** for the binomial variable label in cell A1 of an Excel worksheet.
2. Type **P(X)** for the corresponding probabilities in cell B1.
3. Enter the integers from 0 to 20 in column A, starting at cell A2. Select the Data tab from the toolbar. Then select Data Analysis. Under Analysis Tools, select Random Number Generation and click [OK].
4. In the Random Number Generation dialog box, enter the following:
   a) Number of Variables: **1**
   b) Distribution: Patterned
   c) Parameters: From **0** to **20** in steps of **1,** repeating each number: **1** times and repeating each sequence **1** times
   d) Output range: **A2:A21**
5. Then click [OK].

Random Number Generation Dialog Box



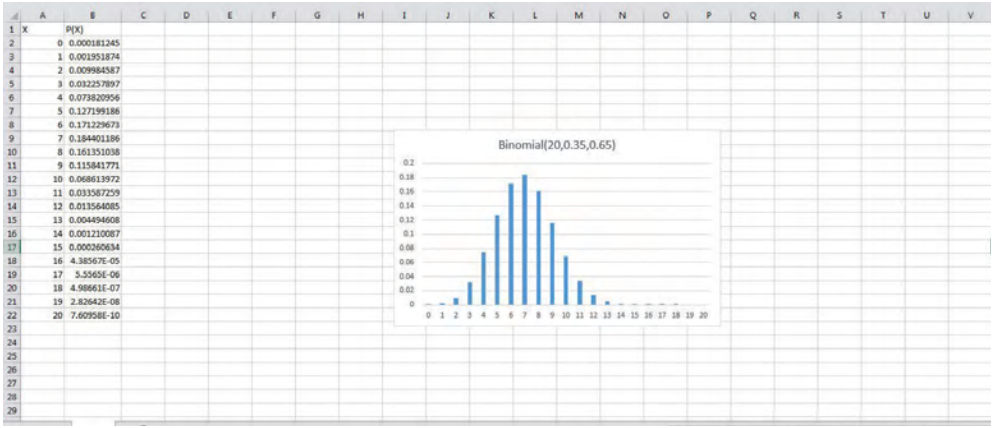
6. To determine the probability corresponding to the first value of the binomial random variable, select cell B2 and type: **=BINOM.DIST(0,20,.35,FALSE).** This will give the probability of obtaining 0 successes in 20 trials of a binomial experiment for which the probability of success is 0.35.
7. Repeat step **6,** changing the first parameter, for each of the values of the random variable from column A.

Note: If you wish to obtain the cumulative probabilities for each of the values in column A, you can type: **=BINOM.DIST(0,20,.35,TRUE)** and repeat for each of the values in column A.

To create the graph:

1. Highlight the probabilities by clicking cell B2 and dragging the mouse down to cell B22. All the cells with probabilities should now be highlighted.
2. Click the *insert* tab from the toolbar and then the button *insert column or bar chart*. (This is the first button on the top row next to recommended charts.)

3. Select the *clustered column* chart (the first column chart under the 2-D Column selections).
4. You will need to edit the horizontal labels to represent the values of X.
   a) Right click anywhere in the chart and choose *select data*.
   b) Click *Edit* under Horizontal (Category) Axis Labels.
   c) Select the values of X by clicking cell A2 and dragging the mouse down to cell A22. All cells with X values should now be highlighted. Click OK.
   d) Click OK to verify these changes overall.
5. To change the title of the chart:
   a) Left-click once on the current title.
   b) Type a new title for the chart, for example, Binomial Distribution (20, .35, .65).



MINITAB
Step by Step

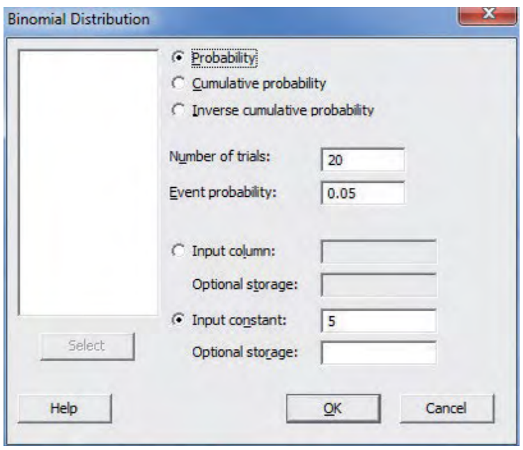
The Binomial Distribution

Calculate a Binomial Probability

From Example 5–20, it is known that 5% of the population is afraid of being alone at night. If a random sample of 20 Americans is selected, what is the probability that exactly 5 of them are afraid?

$$n = 20 \qquad p = 0.05\ (5\%) \qquad \text{and} \qquad X = 5\ (5 \text{ out of } 20)$$

No data need to be entered in the worksheet.



1. Select **Calc>Probability Distributions>Binomial.**
2. Click the option for Probability.
3. Click in the text box for Number of trials:.
4. Type in **20,** then Tab to Event Probability, then type **.05.**
5. Click the option for Input constant, then type in **5.** Leave the text box for Optional storage empty. If the name of a constant such as K1 is entered here, the results are stored but not displayed in the session window.
6. Click [OK]. The results are visible in the session window.

Probability Density Function

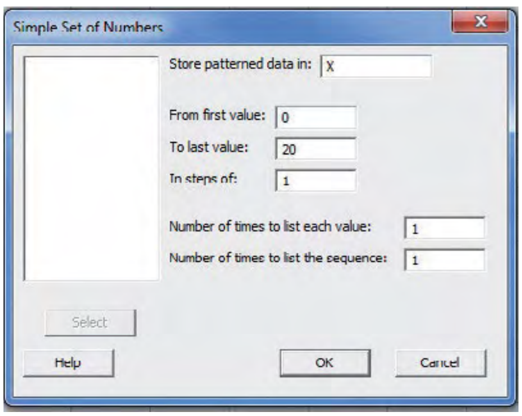
Binomial with n = 20 and p = 0.05

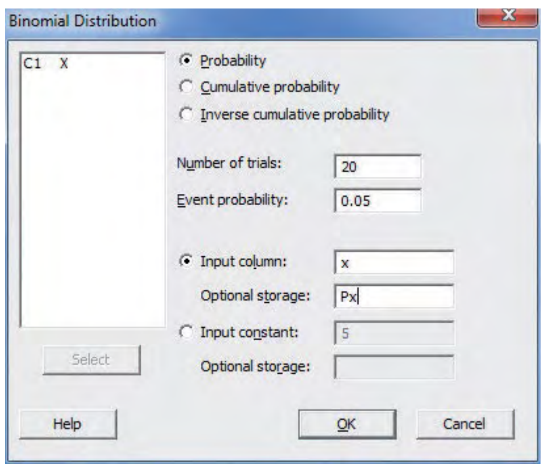
x	P(X = x)
5	0.0022446

Construct a Binomial Distribution

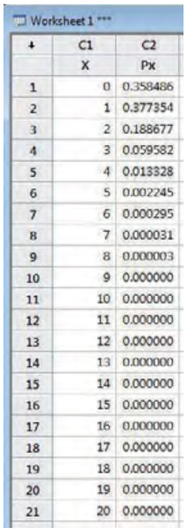
These instructions will use $n = 20$ and $p = 0.05$.

1. Select **Calc>Make Patterned Data>Simple Set of Numbers.**
2. You must enter three items:
   a) Enter **X** in the box for Store patterned data in:. MINITAB will use the first empty column of the active worksheet and name it X.
   b) Press Tab. Enter the value of **0** for the first value. Press Tab.
   c) Enter **20** for the last value. This value should be *n*. In steps of:, the value should be 1.
3. Click [OK].
4. Select **Calc>Probability Distributions>Binomial.**
5. In the dialog box you must enter five items.
   a) Click the button for Probability.
   b) In the box for Number of trials enter **20.**
   c) Enter **.05** in the Event Probability.





   d) Check the button for Input columns, then type the column name, **X,** in the text box.
   e) Click in the box for Optional storage, then type **Px.**



Worksheet 1 ***

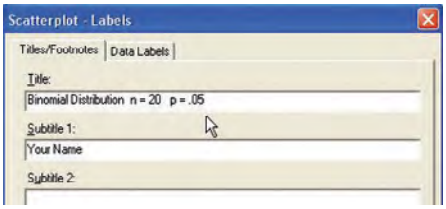
↓	C1	C2
	X	Px
1	0	0.358486
2	1	0.377354
3	2	0.188677
4	3	0.059582
5	4	0.013328
6	5	0.002245
7	6	0.000295
8	7	0.000031
9	8	0.000003
10	9	0.000000
11	10	0.000000
12	11	0.000000
13	12	0.000000
14	13	0.000000
15	14	0.000000
16	15	0.000000
17	16	0.000000
18	17	0.000000
19	18	0.000000
20	19	0.000000
21	20	0.000000

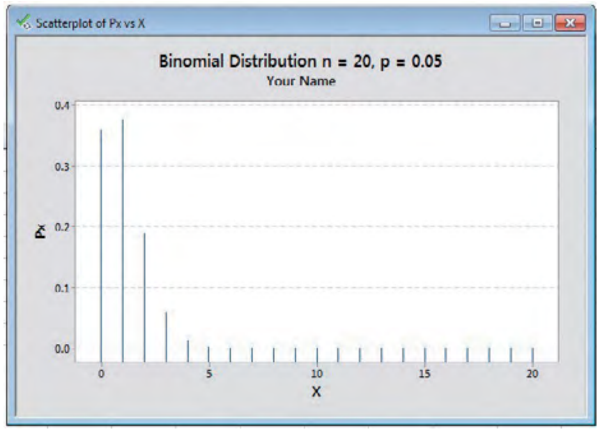
6. Click [OK]. The first available column will be named Px, and the calculated probabilities will be stored in it.
7. To view the completed table, click the worksheet icon on the toolbar . This table matches Table B, the binomial distribution found in Appendix A.

Graph a Binomial Distribution

The table must be available in the worksheet.

1. Select **Graph>Scatterplot,** then Simple.
   a) Double-click on C2 Px for the Y variable and C1 X for the X variable.
   b) Click [Data view], then Project lines, then [OK]. Deselect any other type of display that may be selected in this list.
   c) Click on [Labels], then Title/Footnotes.
   d) Type an appropriate title, such as **Binomial Distribution n = 20, p = .05.**
   e) Press Tab to the Subtitle 1, then type in Your Name then [OK].
   f) Optional: Click [Scales] then [Gridlines], then check the box for Y major ticks.
   g) Click [OK] twice.





The graph will be displayed in a window. Right-click the control box to save, print, or close the graph.

5–4 Other Types of Distributions

OBJECTIVE 5

Find probabilities for outcomes of variables, using the Poisson, hypergeometric, geometric, and multinomial distributions.

In addition to the binomial distribution, other types of distributions are used in statistics. Four of the most commonly used distributions are the multinomial distribution, the Poisson distribution, the hypergeometric distribution, and the geometric distribution. They are described next.

The Multinomial Distribution

Recall that for an experiment to be binomial, two outcomes are required for each trial. But if each trial in an experiment has more than two outcomes, a distribution called the **multinomial distribution** must be used. For example, a survey might require the responses of “approve,” “disapprove,” or “no opinion.” In another situation, a person may have a choice of one of five activities for Friday night, such as a movie, dinner, baseball game, play, or party. Since these situations have more than two possible outcomes for each trial, the binomial distribution cannot be used to compute probabilities.

The multinomial distribution can be used for such situations.

A **multinomial experiment** is a probability experiment that satisfies the following four requirements:

1. There must be a fixed number of trials.
2. Each trial has a specific—but not necessarily the same—number of outcomes.
3. The trials are independent.
4. The probability of a particular outcome remains the same.

Formula for the Multinomial Distribution

If X consists of events $E_1, E_2, E_3, \ldots, E_k$, which have corresponding probabilities $p_1, p_2, p_3, \ldots, p_k$ of occurring, and X_1 is the number of times E_1 will occur, X_2 is the number of times E_2 will occur, X_3 is the number of times E_3 will occur, etc., then the probability that X will occur is

$$P(X) = \frac{n!}{X_1! \cdot X_2! \cdot X_3! \cdots X_k!} \cdot p_1^{X_1} \cdot p_2^{X_2} \cdots p_k^{X_k}$$

where $X_1 + X_2 + X_3 + \cdots + X_k = n$ and $p_1 + p_2 + p_3 + \cdots + p_k = 1$.

EXAMPLE 5–25 Herbicides

It was found that 65% of individuals use herbicides for commercial purposes, 27% of individuals use herbicides for agricultural purposes, and 8% of individuals use herbicides for home and garden purposes. Of 5 people who said that they used herbicides, find the probability that 3 used them for commercial purposes, 1 used them for agriculture purposes, and 1 used them for home or garden purposes.

Source: EPA.

SOLUTION

Let $n = 5$, $X_1 = 3$, $X_2 = 1$, $X_3 = 1$, $p_1 = 0.65$, $p_2 = 0.27$, and $p_3 = 0.08$. Substituting in the formula gives

$$P(X) = \frac{5!}{3! \cdot 1! \cdot 1!} \cdot (0.65)^3(0.27)^1(0.08)^1 \approx 0.119$$

There is a 0.119 probability that if 5 people are selected, 3 will use herbicides for commercial purposes, 1 person will use them for agricultural purposes, and 1 person will use them for home and garden purposes.

Again, note that the multinomial distribution can be used even though replacement is not done, provided that the sample is small in comparison with the population.

EXAMPLE 5–26 Coffee Shop Customers

A small airport coffee shop manager found that the probabilities a customer buys 0, 1, 2, or 3 cups of coffee are 0.3, 0.5, 0.15, and 0.05, respectively. If 8 customers enter the shop, find the probability that 2 will purchase something other than coffee, 4 will purchase 1 cup of coffee, 1 will purchase 2 cups, and 1 will purchase 3 cups.

SOLUTION

Let $n = 8$, $X_1 = 2$, $X_2 = 4$, $X_3 = 1$, and $X_4 = 1$.

$$p_1 = 0.3 \qquad p_2 = 0.5 \qquad p_3 = 0.15 \qquad \text{and} \qquad p_4 = 0.05$$

Then

$$P(X) = \frac{8!}{2!4!1!1!} \cdot (0.3)^2(0.5)^4(0.15)^1(0.05)^1 \approx 0.0354$$

There is a 0.0354 probability that the results will occur as described.

EXAMPLE 5–27 Selecting Colored Balls

A box contains 4 white balls, 3 red balls, and 3 blue balls. A ball is selected at random, and its color is written down. It is replaced each time. Find the probability that if 5 balls are selected, 2 are white, 2 are red, and 1 is blue.

SOLUTION

We know that $n = 5$, $X_1 = 2$, $X_2 = 2$, $X_3 = 1$; $p_1 = \frac{4}{10}$, $p_2 = \frac{3}{10}$, and $p_3 = \frac{3}{10}$; hence,

$$P(X) = \frac{5!}{2!2!1!} \cdot \left(\frac{4}{10}\right)^2\left(\frac{3}{10}\right)^2\left(\frac{3}{10}\right)^1 = \frac{81}{625} = 0.1296$$

There is a 0.1296 probability that the results will occur as described.

Thus, the multinomial distribution is similar to the binomial distribution but has the advantage of allowing you to compute probabilities when there are more than two outcomes for each trial in the experiment. That is, the multinomial distribution is a general distribution, and the binomial distribution is a special case of the multinomial distribution.

Historical Notes

Simeon D. Poisson (1781–1840) formulated the distribution that bears his name. It appears only once in his writings and is only one page long. Mathematicians paid little attention to it until 1907, when a statistician named W. S. Gosset found real applications for it.

The Poisson Distribution

A discrete probability distribution that is useful when n is large and p is small and when the independent variables occur over a period of time is called the **Poisson distribution.** In addition to being used for the stated conditions (that is, n is large, p is small, and the variables occur over a period of time), the Poisson distribution can be used when a density of items is distributed over a given area or volume, such as the number of plants growing per acre or the number of defects in a given length of videotape.

A **Poisson experiment** is a probability experiment that satisfies the following requirements:

1. The random variable X is the number of occurrences of an event over some interval (i.e., length, area, volume, period of time, etc.).
2. The occurrences occur randomly.
3. The occurrences are independent of one another.
4. The average number of occurrences over an interval is known.

Formula for the Poisson Distribution

The probability of X occurrences in an interval of time, volume, area, etc., for a variable where λ (Greek letter lambda) is the mean number of occurrences per unit (time, volume, area, etc.) is

$$P(X; \lambda) = \frac{e^{-\lambda}\lambda^X}{X!} \qquad \text{where } X = 0, 1, 2, \ldots$$

The letter e is a constant approximately equal to 2.7183.

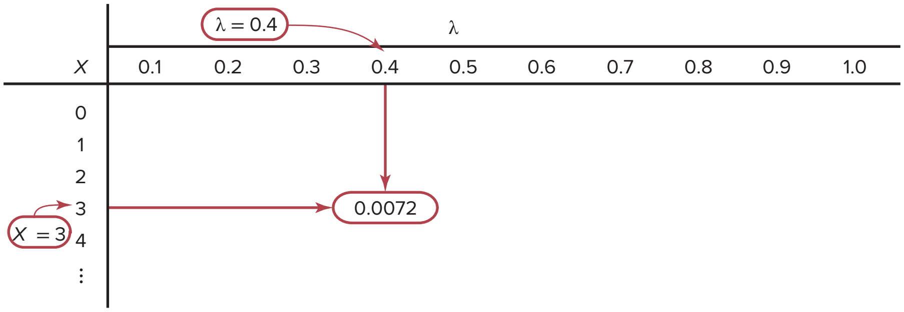


FIGURE 5–4
Using Table C

Round the answers to four decimal places.

EXAMPLE 5–28 Typographical Errors

If there are 200 typographical errors randomly distributed in a 500-page manuscript, find the probability that a given page contains exactly 3 errors.

SOLUTION

First, find the mean number λ of errors. Since there are 200 errors distributed over 500 pages, each page has an average of

$$\lambda = \frac{200}{500} = \frac{2}{5} = 0.4$$

or 0.4 error per page. Since $X = 3$, substituting into the formula yields

$$P(X; \lambda) = \frac{e^{-\lambda}\lambda^X}{X!} = \frac{(2.7183)^{-0.4}(0.4)^3}{3!} \approx 0.0072$$

Thus, there is less than a 1% chance that any given page will contain exactly 3 errors.

Since the mathematics involved in computing Poisson probabilities is somewhat complicated, tables have been compiled for these probabilities. Table C in Appendix A gives P for various values for λ and X.

In Example 5–28, where X is 3 and λ is 0.4, the table gives the value 0.0072 for the probability. See Figure 5–4.

EXAMPLE 5–29 Toll-Free Telephone Calls

A sales firm receives, on average, 3 calls per hour on its toll-free number. For any given hour, find the probability that it will receive the following.

a. At most 3 calls *b.* At least 3 calls *c.* 5 or more calls

SOLUTION

a. "At most 3 calls" means 0, 1, 2, or 3 calls. Hence,

$$\begin{aligned} &P(0; 3) + P(1; 3) + P(2; 3) + P(3; 3) \\ &\quad = 0.0498 + 0.1494 + 0.2240 + 0.2240 \\ &\quad = 0.6472 \end{aligned}$$

b. "At least 3 calls" means 3 or more calls. It is easier to find the probability of 0, 1, and 2 calls and then subtract this answer from 1 to get the probability of at least 3 calls.

$$P(0; 3) + P(1; 3) + P(2; 3) = 0.0498 + 0.1494 + 0.2240 = 0.4232$$

and

$$1 - 0.4232 = 0.5768$$

c. For the probability of 5 or more calls, it is easier to find the probability of getting 0, 1, 2, 3, or 4 calls and subtract this answer from 1. Hence,

$$\begin{aligned} &P(0; 3) + P(1; 3) + P(2; 3) + P(3; 3) + P(4; 3) \\ &\quad = 0.0498 + 0.1494 + 0.2240 + 0.2240 + 0.1680 \\ &\quad = 0.8152 \end{aligned}$$

and

$$1 - 0.8152 = 0.1848$$

Thus, for the events described, the part *a* event is most likely to occur, and the part *c* event is least likely to occur.

The Poisson distribution can also be used to approximate the binomial distribution when the expected value $\lambda = n \cdot p$ is less than 5, as shown in Example 5–30. (The same is true when $n \cdot q < 5$.)

EXAMPLE 5–30 Left-Handed People

If approximately 2% of the people in a room of 200 people are left-handed, find the probability that exactly 5 people there are left-handed.

SOLUTION

Since $\lambda = n \cdot p$, then $\lambda = (200)(0.02) = 4$. Hence,

$$P(X; \lambda) = \frac{(2.7183)^{-4}(4)^5}{5!} \approx 0.1563$$

which is verified by the formula ${}_{200}C_5(0.02)^5(0.98)^{195} \approx 0.1579$. The difference between the two answers is based on the fact that the Poisson distribution is an approximation and rounding has been used.

The Hypergeometric Distribution

When sampling is done *without* replacement, the binomial distribution does not give exact probabilities, since the trials are not independent. The smaller the size of the population, the less accurate the binomial probabilities will be.

For example, suppose a committee of 4 people is to be selected from 7 women and 5 men. What is the probability that the committee will consist of 3 women and 1 man?

To solve this problem, you must find the number of ways a committee of 3 women and 1 man can be selected from 7 women and 5 men. This answer can be found by using combinations; it is

$${}_7C_3 \cdot {}_5C_1 = 35 \cdot 5 = 175$$

Next, find the total number of ways a committee of 4 people can be selected from 12 people. Again, by the use of combinations, the answer is

$$_{12}C_4 = 495$$

Finally, the probability of getting a committee of 3 women and 1 man from 7 women and 5 men is

$$P(X) = \frac{175}{495} = \frac{35}{99}$$

The results of the problem can be generalized by using a special probability distribution called the hypergeometric distribution. The **hypergeometric distribution** is a distribution of a variable that has two outcomes when sampling is done without replacement.

A **hypergeometric experiment** is a probability experiment that satisfies the following requirements:

1. There are a fixed number of trials.
2. There are two outcomes, and they can be classified as success or failure.
3. The sample is selected without replacement.

The probabilities for the hypergeometric distribution can be calculated by using the formula given next.

Formula for the Hypergeometric Distribution

Given a population with only two types of objects (females and males, defective and nondefective, successes and failures, etc.), such that there are a items of one kind and b items of another kind and $a + b$ equals the total population, the probability $P(X)$ of selecting without replacement a sample of size n with X items of type a and $n - X$ items of type b is

$$P(X) = \frac{_aC_X \cdot {_bC_{n-X}}}{_{a+b}C_n}$$

The basis of the formula is that there are $_aC_X$ ways of selecting the first type of items, $_bC_{n-X}$ ways of selecting the second type of items, and $_{a+b}C_n$ ways of selecting n items from the entire population.

EXAMPLE 5–31 Assistant Manager Applicants

Ten people apply for a job as assistant manager of a restaurant. Five have completed college and five have not. If the manager selects 3 applicants at random, find the probability that all 3 are college graduates.

SOLUTION

Assigning the values to the variables gives

$$a = 5 \text{ college graduates} \qquad n = 3$$
$$b = 5 \text{ nongraduates} \qquad X = 3$$

and $n - X = 0$. Substituting in the formula gives

$$P(X) = \frac{_5C_3 \cdot {_5C_0}}{_{10}C_3} = \frac{10}{120} = \frac{1}{12} \approx 0.083$$

There is a 0.083 probability that all 3 applicants will be college graduates.

EXAMPLE 5–32 House Insurance

A recent study found that 2 out of every 10 houses in a neighborhood have no insurance. If 5 houses are selected from 10 houses, find the probability that exactly 1 will be uninsured.

SOLUTION

In this example, $a = 2$, $b = 8$, $n = 5$, $X = 1$, and $n - X = 4$.

$$P(X) = \frac{{}_2C_1 \cdot {}_8C_4}{{}_{10}C_5} = \frac{2 \cdot 70}{252} = \frac{140}{252} = \frac{5}{9} \approx 0.556$$

There is a 0.556 probability that out of 5 houses, 1 house will be uninsured.

In many situations where objects are manufactured and shipped to a company, the company selects a few items and tests them to see whether they are satisfactory or defective. If a certain percentage is defective, the company then can refuse the whole shipment. This procedure saves the time and cost of testing every single item. To make the judgment about whether to accept or reject the whole shipment based on a small sample of tests, the company must know the probability of getting a specific number of defective items. To calculate the probability, the company uses the hypergeometric distribution.

EXAMPLE 5–33 Defective Compressor Tanks

A lot of 12 compressor tanks is checked to see whether there are any defective tanks. Three tanks are checked for leaks. If 1 or more of the 3 is defective, the lot is rejected. Find the probability that the lot will be rejected if there are actually 3 defective tanks in the lot.

SOLUTION

Since the lot is rejected if at least 1 tank is found to be defective, it is necessary to find the probability that none are defective and subtract this probability from 1.

Here, $a = 3$, $b = 9$, $n = 3$, and $X = 0$; so

$$P(X) = \frac{{}_3C_0 \cdot {}_9C_3}{{}_{12}C_3} = \frac{1 \cdot 84}{220} \approx 0.382$$

Hence,

$$P(\text{at least 1 defective}) = 1 - P(\text{no defectives}) = 1 - 0.382 = 0.618$$

There is a 0.618 or 61.8%, probability that the lot will be rejected when 3 of the 12 tanks are defective.

The Geometric Distribution

Another useful distribution is called the *geometric distribution.* This distribution can be used when we have an experiment that has two outcomes and is repeated until a successful outcome is obtained. For example, we could flip a coin until a head is obtained, or we could roll a die until we get a 6. In these cases, our successes would come on the nth trial. The geometric probability distribution tells us when the success is likely to occur.

A **geometric experiment** is a probability experiment if it satisfies the following requirements:

1. Each trial has two outcomes that can be either success or failure.
2. The outcomes are independent of each other.
3. The probability of a success is the same for each trial.
4. The experiment continues until a success is obtained.

Formula for the Geometric Distribution

If p is the probability of a success on each trial of a binomial experiment and n is the number of the trial at which the first success occurs, then the probability of getting the first success on the nth trial is

$$P(n) = p(1 - p)^{n-1}$$

where $n = 1, 2, 3, \ldots$.

EXAMPLE 5–34 Rolling a Die

A die is rolled repeatedly. Find the probability of getting the first 2 on the third roll.

SOLUTION

To get a 2 on the third roll, the first two rolls must be any other number except a 2; hence, the probability is P(not 2, not 2, 2) is

$$\frac{5}{6} \cdot \frac{5}{6} \cdot \frac{1}{6} = \frac{25}{216}$$

Now by using the formula you get the same results:

$$\begin{aligned} P(n) &= p(1 - p)^{n-1} \\ &= \frac{1}{6}\left(1 - \frac{1}{6}\right)^2 \\ &= \frac{1}{6}\left(\frac{5}{6}\right)^2 \\ &= \frac{25}{216} \end{aligned}$$

Hence, the probability of getting the first 2 on the third roll is $\frac{25}{216}$.

EXAMPLE 5–35 Blood Types

In the United States, approximately 42% of people have type A blood. If 4 people are selected at random, find the probability that the fourth person is the first one selected with type A blood.

SOLUTION

Let $p = 0.42$ and $n = 4$.

$$\begin{aligned} P(n) &= p(1 - p)^{n-1} \\ P(4) &= (0.42)(1 - 0.42)^{4-1} \\ &= (0.42)(0.58)^3 \\ &\approx 0.0819 \approx 0.082 \end{aligned}$$

There is a 0.082 probability that the fourth person selected will be the first one to have type A blood.

A summary of the discrete distributions used in this chapter is shown in Table 5–1.

Interesting Fact

An IBM supercomputer set a world record in 2008 by performing 1.026 quadrillion calculations in 1 second.

TABLE 5–1 Summary of Discrete Distributions

1. Binomial distribution

$$P(X) = \frac{n!}{(n-X)!X!} \cdot p^X \cdot q^{n-X}$$

$$\mu = n \cdot p \qquad \sigma = \sqrt{n \cdot p \cdot q}$$

It is used when there are only two outcomes for a fixed number of independent trials and the probability for each success remains the same for each trial.

2. Multinomial distribution

$$P(X) = \frac{n!}{X_1! \cdot X_2! \cdot X_3! \cdots X_k!} \cdot p_1^{X_1} \cdot p_2^{X_2} \cdots p_k^{X_k}$$

where

$$X_1 + X_2 + X_3 + \cdots + X_k = n \qquad \text{and} \qquad p_1 + p_2 + p_3 + \cdots + p_k = 1$$

It is used when the distribution has more than two outcomes, the probabilities for each trial remain constant, outcomes are independent, and there are a fixed number of trials.

3. Poisson distribution

$$P(X; \lambda) = \frac{e^{-\lambda}\lambda^X}{X!} \qquad \text{where } X = 0, 1, 2, \ldots$$

It is used when n is large and p is small, and the independent variable occurs over a period of time, or a density of items is distributed over a given area or volume.

4. Hypergeometric distribution

$$P(X) = \frac{{}_aC_X \cdot {}_bC_{n-X}}{{}_{a+b}C_n}$$

It is used when there are two outcomes and sampling is done without replacement.

5. Geometric distribution

$$P(n) = p(1-p)^{n-1} \qquad \text{where } n = 1, 2, 3, \ldots$$

It is used when there are two outcomes and we are interested in the probability that the first success occurs on the nth trial.

Applying the Concepts 5–4

Rockets and Targets

During the latter days of World War II, the Germans developed flying rocket bombs. These bombs were used to attack London. Allied military intelligence didn't know whether these bombs were fired at random or had a sophisticated aiming device. To determine the answer, they used the Poisson distribution.

To assess the accuracy of these bombs, London was divided into 576 square regions. Each region was $\frac{1}{4}$ square kilometer in area. They then compared the number of actual hits with the theoretical number of hits by using the Poisson distribution. If the values in both distributions were close, then they would conclude that the rockets were fired at random. The actual distribution is as follows:

Hits	0	1	2	3	4	5
Regions	229	211	93	35	7	1

1. Using the Poisson distribution, find the theoretical values for each number of hits. In this case, the number of bombs was 535, and the number of regions was 576. So

$$\lambda = \frac{535}{576} \approx 0.929$$

For 3 hits,

$$P(X) = \frac{e^{-\lambda} \cdot \lambda^X}{X!}$$
$$= \frac{(2.7183)^{-0.929}(0.929)^3}{3!} \approx 0.0528$$

Hence, the number of hits is (0.0528)(576) = 30.4128.
Complete the table for the other number of hits.

Hits	0	1	2	3	4	5
Regions				30.4		

2. Write a brief statement comparing the two distributions.
3. Based on your answer to question 2, can you conclude that the rockets were fired at random?

See page 310 for the answer.

Exercises 5–4

1. Use the multinomial formula and find the probabilities for each.

a. $n = 6, X_1 = 3, X_2 = 2, X_3 = 1, p_1 = 0.5, p_2 = 0.3, p_3 = 0.2$

b. $n = 5, X_1 = 1, X_2 = 2, X_3 = 2, p_1 = 0.3, p_2 = 0.6, p_3 = 0.1$

c. $n = 4, X_1 = 1, X_2 = 1, X_3 = 2, p_1 = 0.8, p_2 = 0.1, p_3 = 0.1$

2. Use the multinomial formula and find the probabilities for each.

a. $n = 3, X_1 = 1, X_2 = 1, X_3 = 1, p_1 = 0.5, p_2 = 0.3, p_3 = 0.2$

b. $n = 5, X_1 = 1, X_2 = 3, X_3 = 1, p_1 = 0.7, p_2 = 0.2, p_3 = 0.1$

c. $n = 7, X_1 = 2, X_2 = 3, X_3 = 2, p_1 = 0.4, p_2 = 0.5, p_3 = 0.1$

3. M&M's Color Distribution According to the manufacturer, M&M's are produced and distributed in the following proportions: 13% brown, 13% red, 14% yellow, 16% green, 20% orange, and 24% blue. In a random sample of 12 M&M's, what is the probability of having 2 of each color?

4. Truck Inspection Violations The probabilities are 0.50, 0.40, and 0.10 that a trailer truck will have no violations, 1 violation, or 2 or more violations when it is given a safety inspection by state police. If 5 trailer trucks are inspected, find the probability that 3 will have no violations, 1 will have 1 violation, and 1 will have 2 or more violations.

5. Drug Prescriptions The probability that a person has 1, 2, 3, or 4 prescriptions when he or she enters a pharmacy is 0.5, 0.3, 0.15, or 0.05. For a sample of 10 people who enter the pharmacy, find the probability that 3 will have one prescription filled, 3 will have two prescriptions filled, 2 will have three perscriptions filled, and 2 will have four prescriptions filled.

6. Mendel's Theory According to Mendel's theory, if tall and colorful plants are crossed with short and colorless plants, the corresponding probabilities are $\frac{9}{16}$, $\frac{3}{16}$, $\frac{3}{16}$, and $\frac{1}{16}$ for tall and colorful, tall and colorless, short and colorful, and short and colorless, respectively. If 8 plants are selected, find the probability that 1 will be tall and colorful, 3 will be tall and colorless, 3 will be short and colorful, and 1 will be short and colorless.

7. Find each probability $P(X; \lambda)$, using Table C in Appendix A.

a. $P(6;4)$

b. $P(2;5)$

c. $P(7;3)$

8. Find each probability $P(X; \lambda)$ using Table C in Appendix A.

a. $P(10; 7)$

b. $P(9; 8)$

c. $P(3; 4)$

9. Study of Robberies A recent study of robberies for a certain geographic region showed an average of 1 robbery per 20,000 people. In a city of 80,000 people, find the probability of the following.

a. 0 robberies

b. 1 robbery

c. 2 robberies

d. 3 or more robberies

10. Misprints on Manuscript Pages In a 400-page manuscript, there are 200 randomly distributed misprints. If a page is selected, find the probability that it has 1 misprint.

11. Colors of Flowers A nursery provides red impatiens for commercial landscaping. If 5% are variegated instead of pure red, find the probability that in an order for 200 plants, exactly 14 are variegated.

12. Mail Ordering A mail-order company receives an average of 5 orders per 500 solicitations. If it sends out 100 advertisements, find the probability of receiving at least 2 orders.

13. Company Mailing Of a company's mailings 1.5% are returned because of incorrect or incomplete addresses. In a mailing of 200 pieces, find the probability that none will be returned.

14. Emission Inspection Failures If 3% of all cars fail the emissions inspection, find the probability that in a sample of 90 cars, 3 will fail. Use the Poisson approximation.

15. Phone Inquiries The average number of phone inquiries per day at the poison control center is 4. Find the probability it will receive 5 calls on a given day. Use the Poisson approximation.

16. Defective Calculators In a batch of 2000 calculators, there are, on average, 8 defective ones. If a random sample of 150 is selected, find the probability of 5 defective ones.

17. School Newspaper Staff A school newspaper staff is comprised of 5 seniors, 4 juniors, 5 sophomores, and 7 freshmen. If 4 staff members are chosen at random for a publicity photo, what is the probability that there will be 1 student from each class?

18. Missing Pages from Books A bookstore owner examines 5 books from each lot of 25 to check for missing pages. If he finds at least 2 books with missing pages, the entire lot is returned. If, indeed, there are 5 books with missing pages, find the probability that the lot will be returned.

19. Job Applicants Twelve people apply for a teaching position in mathematics at a local college. Six have a PhD and six have a master's degree. If the department chairperson selects three applicants at random for an interview, find the probability that all three have a PhD.

20. Defective Computer Keyboards A shipment of 24 computer keyboards is rejected if 4 are checked for defects and at least 1 is found to be defective. Find the probability that the shipment will be returned if there are actually 6 defective keyboards.

21. Defective Electronics A shipment of 24 smartphones is rejected if 3 are checked for defects and at least 1 is found to be defective. Find the probability that the shipment will be returned if there are actually 6 smartphones that are defective.

22. Job Applications Ten people apply for a job at Computer Warehouse. Five are college graduates and five are not. If the manager selects 3 applicants at random, find the probability that all 3 are college graduates.

23. Auto Repair Insurance A person calls people to ask if they would like to extend their automobile insurance beyond the normal 3 years. The probability that the respondent says yes is about 33%. If she calls 12 people, find the probability that the first person to say yes will occur with the fourth customer.

24. Winning a Prize A soda pop manufacturer runs a contest and places a winning bottle cap on every sixth bottle. If a person buys the soda pop, find the probability that the person will (*a*) win on his first purchase, (*b*) win on his third purchase, or (*c*) not win on any of his first five purchases.

25. Shooting an Arrow Mark shoots arrows at a target and hits the bull's-eye about 40% of the time. Find the probability that he hits the bull's-eye on the third shot.

26. Amusement Park Game At an amusement park basketball game, the player gets 3 throws for $1. If the player makes a basket, the player wins a prize. Mary makes about 80% of her shots. Find the probability that Mary wins a prize on her third shot.

Extending the Concepts

Another type of problem that can be solved uses what is called the *negative binomial distribution,* which is a generalization of the binomial distribution. In this case, it tells the average number of trials needed to get k successes of a binomial experiment. The formula is

$$\mu = \frac{k}{p}$$

where k = the number of successes
p = the probability of a success

Use this formula for Exercises 27–30.

27. Drawing Cards A card is randomly drawn from a deck of cards and then replaced. The process continues until 3 clubs are obtained. Find the average number of trials needed to get 3 clubs.

28. Rolling an 8-Sided Die An 8-sided die is rolled. The sides are numbered 1 through 8. Find the average number of rolls it takes to get two 5s.

29. Drawing Cards Cards are drawn at random from a deck and replaced after each draw. Find the average number of cards that would be drawn to get 4 face cards.

30. Blood Type About 4% of the citizens of the United States have type AB blood. If an agency needed type AB blood and donors came in at random, find the average number of donors that would be needed to get a person with type AB blood.

The mean of a geometric distribution is $\mu = 1/p$, and the standard deviation is $\sigma = \sqrt{q/p^2}$, where p = the probability of the outcome and $q = 1 - p$. Use these formulas for Exercises 31–34.

31. Shower or Bath Preferences It is estimated that 4 out of 5 men prefer showers to baths. Find the mean and standard deviation for the distribution of men who prefer showers to baths.

32. Lessons Outside of School About 2 out of every 3 children take some kind of lessons outside of school. These lessons include music, art, and sports. Find the mean and standard deviation of the distribution of the number of children who take lessons outside of school.

33. Teachers and Summer Vacations One in five teachers stated that he or she became a teacher because of the long summer vacations. Find the mean and standard deviation for the distribution of teachers who say they became teachers because of the long summer vacation.

34. Work versus Conscience One worker in four in America admits that she or he has to do some things at work that go against her or his consciences. Find the mean and standard deviation for the distribution of workers who admit to having to do some things at work that go against their consciences.

Technology Step by Step

TI-84 Plus Step by Step

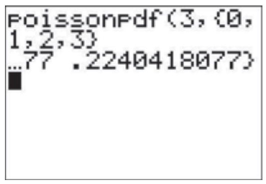
Poisson Random Variables

To find the probability for a Poisson random variable:
Press **2nd [DISTR]** then **C (ALPHA PRGM)** for poissonpdf

Note the form is different from that used in text, $P(X; \lambda)$. And on some calculators (as with binomial cdf and pdf) you will be presented with a menu for inputting X and λ.

Example: $\lambda = 0.4$, $X = 3$ (Example 5–28 from the text)
poissonpdf(.4, 3)

Example: $\lambda = 3$, $X = 0, 1, 2, 3$ (Example 5–29*a* from the text)
poissonpdf (3, {0, 1, 2, 3})
The calculator will display the probabilities in a list. Use the arrow keys to view the entire display.



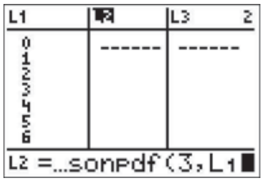
To find the cumulative probability for a Poisson random variable:
Press **2nd [DISTR]** then **D (ALPHA VARS)** for poissoncdf (Note: On the TI-84 Plus use D.)
The form is poissoncdf(λ, X). This will calculate the cumulative probability for values from 0 to X.

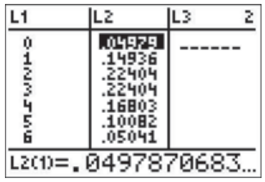
Example: $\lambda = 3$, $X = 0, 1, 2, 3$ (Example 5–29*a* from the text)
poissoncdf(3, 3)

To construct a Poisson probability table:

1. Enter the X values 0 through a large possible value of X into L_1.
2. Move the cursor to the top of the L_2 column so that L_2 is highlighted.
3. Enter the command poissonpdf (λ, L_1) then press **ENTER.**

Example: $\lambda = 3$, $X = 0, 1, 2, 3, \ldots, 10$ (Example 5–29 from the text)





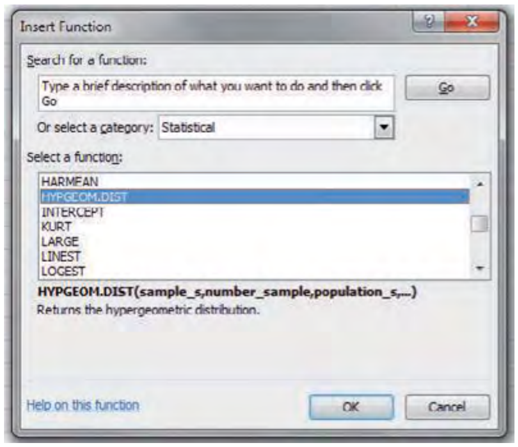
Other Discrete Distributions

Excel can be used to calculate probabilities (and cumulative probabilities). The built-in discrete probability distribution functions available in Excel include the hypergeometric, Poisson, and geometric.

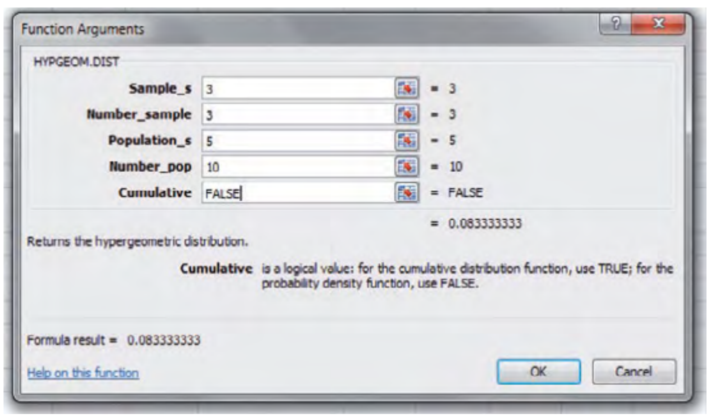
Calculating a Hypergeometric Probability

We will use Excel to calculate the probability from Example 5–31.

1. Select the Insert Function icon from the Toolbar.
2. Select the Statistical function category from the list of available categories.
3. Select the HYPGEOM.DIST function from the function list. The Function Arguments dialog box will appear.



4. Type 3 for Sample_s, the number of successes in the sample.
5. Type 3 for Number_sample, the size of the sample.
6. Type 5 for Population_s, the number of successes in the population.
7. Type 10 for Number_pop, the size of the population.
8. Type FALSE for Cumulative, since the probability to be calculated is for a single event.
9. Click OK.

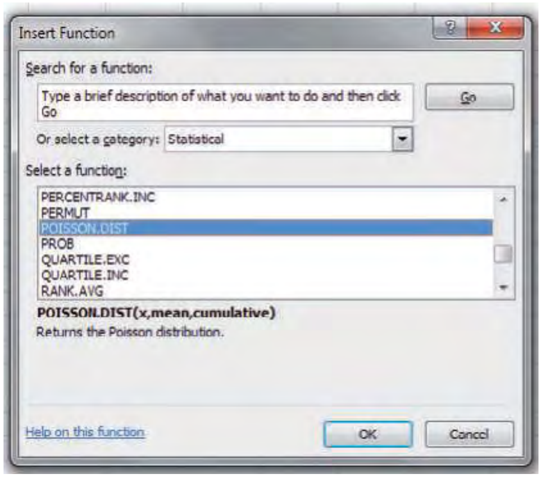


Note: If you are calculating a cumulative probability, you should type TRUE for Cumulative.

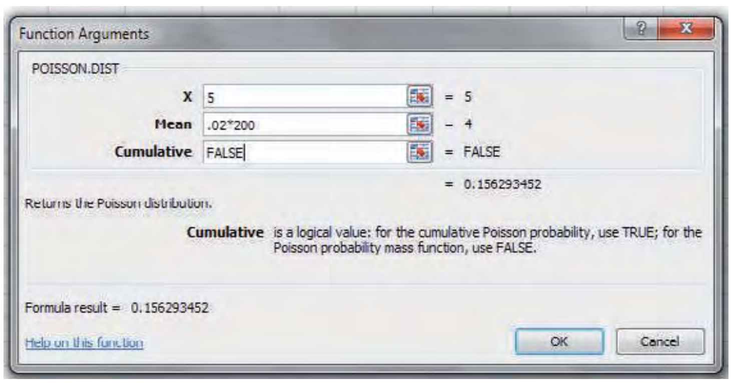
Calculating a Poisson Probability

We will use Excel to calculate the probability from Example 5–30

1. Select the Insert Function Icon from the Toolbar.
2. Select the Statistical function category from the list of available categories.
3. Select the POISSON.DIST function from the function list. The Function Arguments dialog box will appear.



4. Type 5 for *X,* the number of occurrences.
5. Type .02*200 or 4 for the Mean.
6. Type FALSE for Cumulative, since the probability to be calculated is for a single event.
7. Click OK.



Calculating a Geometric Probability

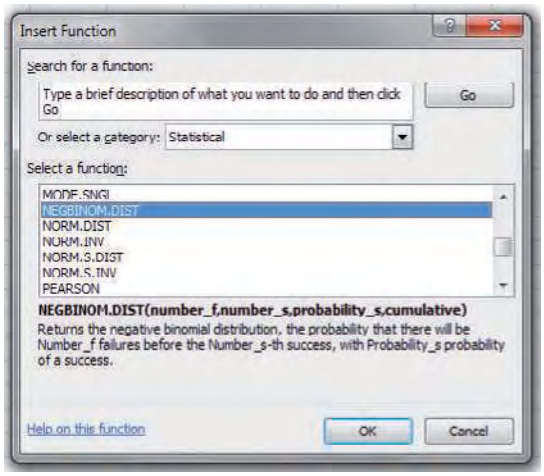
We will use Excel to calculate the probability from Example 5–35.

Note: Excel does not have a built-in Geometric Probability Distribution function. We must use the built-in Negative Binomial Distribution function—which gives the probability that there will be a certain number of failures until a certain number of successes occur—to calculate probabilities for the Geometric Distribution. The Geometric Distribution is a special case of the Negative Binomial for which the threshold number of successes is 1.

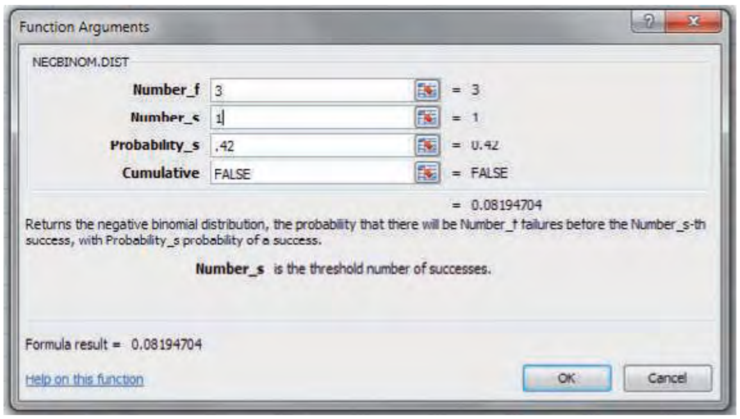
Select the Insert Function Icon from the Toolbar.

1. Select the Statistical function category from the list of available categories.

2. Select the NEGBINOM.DIST function from the function list. The Function Arguments dialog box will appear.



3. When the NEGBINOM.DIST Function Arguments box appears, type 3 for Number_f, the number of failures (until the first success).
4. Type 1 for Number_s, the threshold number of successes.
5. Type .42 for Probability_s, the probability of a success.
6. Type FALSE for cumulative.
7. Click OK.



Summary

- A discrete probability distribution consists of the values a random variable can assume and the corresponding probabilities of these values. There are two requirements of a probability distribution: the sum of the probabilities of the events must equal 1, and the probability of any single event must be a number from 0 to 1. Probability distributions can be graphed. (5–1)
- The mean, variance, and standard deviation of a probability distribution can be found. The expected value of a discrete random variable of a probability distribution can also be found. This is basically a measure of the average. (5–2)
- A binomial experiment has four requirements. There must be a fixed number of trials. Each trial can have only two outcomes. The outcomes are independent of each other, and the probability of a success must remain the same for each trial. The probabilities of the outcomes can be found by using the binomial formula or the binomial table. (5–3)
- In addition to the binomial distribution, there are some other commonly used probability distributions. They are the multinomial distribution, the Poisson distribution, the hypergeometric distribution, and the geometric distribution. (5–4)

Important Terms

binomial distribution 276
binomial experiment 275
discrete probability distribution 259
expected value 269
geometric distribution 295
geometric experiment 296
hypergeometric distribution 294
hypergeometric experiment 294
multinomial distribution 289
multinomial experiment 290
Poisson distribution 291
Poisson experiment 291
random variable 258

Important Formulas

Formula for the mean of a probability distribution:

$$\mu = \Sigma X \cdot P(X)$$

Formulas for the variance and standard deviation of a probability distribution:

$$\sigma^2 = \Sigma[X^2 \cdot P(X)] - \mu^2$$

$$\sigma = \sqrt{\Sigma[X^2 \cdot P(X)] - \mu^2}$$

Formula for expected value:

$$E(X) = \Sigma X \cdot P(X)$$

Binomial probability formula:

$$P(X) = \frac{n!}{(n-X)!X!} \cdot p^X \cdot q^{n-X} \quad \text{where } X = 0, 1, 2, 3, \ldots, n$$

Formula for the mean of the binomial distribution:

$$\mu = n \cdot p$$

Formulas for the variance and standard deviation of the binomial distribution:

$$\sigma^2 = n \cdot p \cdot q \qquad \sigma = \sqrt{n \cdot p \cdot q}$$

Formula for the multinomial distribution:

$$P(X) = \frac{n!}{X_1! \cdot X_2! \cdot X_3! \cdots X_k!} \cdot p_1^{X_1} \cdot p_2^{X_2} \cdots p_k^{X_k}$$

(The X's sum to n and the p's sum to 1.)

Formula for the Poisson distribution:

$$P(X; \lambda) = \frac{e^{-\lambda}\lambda^X}{X!} \qquad \text{where } X = 0, 1, 2, \ldots$$

Formula for the hypergeometric distribution:

$$P(X) = \frac{{}_aC_X \cdot {}_bC_{n-X}}{{}_{a+b}C_n}$$

Formula for the geometric distribution:

$$P(n) = p(1-p)^{n-1} \qquad \text{where } n = 1, 2, 3, \ldots$$

Review Exercises

Section 5–1

For Exercises 1 through 3, determine whether the distribution represents a probability distribution. If it does not, state why.

1.

X	1	2	3	4	5
$P(X)$	$\frac{1}{10}$	$\frac{3}{10}$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{3}{10}$

2.

X	5	10	15
$P(X)$	0.3	0.4	0.1

3.

X	1	4	9	16
$P(X)$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$

4. Emergency Calls The number of emergency calls that a local police department receives per 24-hour period is distributed as shown here. Construct a graph for the data.

Number of calls X	10	11	12	13	14
Probability $P(X)$	0.02	0.12	0.40	0.31	0.15

5. Credit Cards A large retail company encourages its employees to get customers to apply for the store credit card. Below is the distribution for the number of credit card applications received per employee for an 8-hour shift.

X	0	1	2	3	4	5
$P(X)$	0.27	0.28	0.20	0.15	0.08	0.02

a. What is the probability that an employee will get 2 or 3 applications during any given shift?

b. Find the mean, variance, and standard deviation for this probability distribution.

6. Coins in a Box A box contains 5 pennies, 3 dimes, 1 quarter, and 1 half-dollar. A coin is drawn at random. Construct a probability distribution and draw a graph for the data.

7. Shoe Purchases At Shoe World, the manager finds the probability that a woman will buy 1, 2, 3, or 4 pairs of shoes is shown. Construct a graph for the distribution.

Number X	1	2	3	4
Probability $P(X)$	0.63	0.27	0.08	0.02

Section 5–2

8. Customers in a Bank A bank has a drive-through service. The number of customers arriving during a 15-minute period is distributed as shown. Find the mean, variance, and standard deviation for the distribution.

Number of customers X	0	1	2	3	4
Probability $P(X)$	0.12	0.20	0.31	0.25	0.12

9. Arrivals at an Airport At a small rural airport, the number of arrivals per hour during the day has the distribution shown. Find the mean, variance, and standard deviation for the data.

Number X	5	6	7	8	9	10
Probability $P(X)$	0.14	0.21	0.24	0.18	0.16	0.07

10. Cans of Paint Purchased During a recent paint sale at Corner Hardware, the number of cans of paint purchased was distributed as shown. Find the mean, variance, and standard deviation of the distribution.

Number of cans X	1	2	3	4	5
Probability $P(X)$	0.42	0.27	0.15	0.10	0.06

11. Phone Customers A phone service center keeps track of the number of customers it services for a random 1-hour period during the day. The distribution is shown.

Number of calls X	0	1	2	3	4
$P(X)$	0.15	0.43	0.32	0.06	0.04

Find the mean, variance, and standard deviation for the data. Based on the results of the data, how many service representatives should the store employ?

12. Outdoor Regatta A producer plans an outdoor regatta for May 3. The cost of the regatta is \$8000. This includes advertising, security, printing tickets, entertainment, etc. The producer plans to make \$15,000 profit if all goes well. However, if it rains, the regatta will have to be canceled. According to the weather report, the probability of rain is 0.3. Find the producer's expected profit.

13. Card Game A game is set up as follows: All the diamonds are removed from a deck of cards, and these 13 cards are placed in a bag. The cards are mixed up, and then one card is chosen at random (and then replaced). The player wins according to the following rules.

If the ace is drawn, the player loses \$20.
If a face card is drawn, the player wins \$10.
If any other card (2–10) is drawn, the player wins \$2.

How much should be charged to play this game in order for it to be fair?

14. Card Game Using Exercise 13, how much should be charged if instead of winning \$2 for drawing a 2–10, the player wins the amount shown on the card in dollars?

Section 5–3

15. Let x be a binomial random variable with $n = 12$ and $p = 0.3$. Find the following:

a. $P(X = 8)$
b. $P(X < 5)$
c. $P(X \geq 10)$
d. $P(4 < X \leq 9)$

16. Internet Access via Cell Phone In a retirement community, 14% of cell phone users use their cell phones to access the Internet. In a random sample of 10 cell phone users, what is the probability that exactly 2 have used their phones to access the Internet? More than 2?

17. Self-Driving Automobile Fifty-eight percent of people surveyed said that they would take a ride in a fully self-driving automobile. Find the mean, variance, and standard deviation of the number of people who would agree to ride in the self-driving automobile if 250 people were asked.

Source: World Economic Forum and Boston Consulting Group Survey.

18. Flu Shots It has been reported that 63% of adults aged 65 and over got their flu shots last year. In a random sample of 300 adults aged 65 and over, find the mean, variance, and standard deviation for the number who got their flu shots.

Source: U.S. Centers for Disease Control and Prevention.

19. U.S. Police Chiefs and the Death Penalty The chance that a U.S. police chief believes the death penalty "significantly reduces the number of homicides" is 1 in 4. If a random sample of 8 police chiefs is selected, find the probability that at most 3 believe that the death penalty significantly reduces the number of homicides.

Source: Harper's Index.

20. Household Wood Burning *American Energy Review* reported that 27% of American households burn wood. If a random sample of 500 American households is selected, find the mean, variance, and standard deviation of the number of households that burn wood.

Source: 100% American by Daniel Evan Weiss.

21. Pizza for Breakfast Three out of four American adults under age 35 have eaten pizza for breakfast. If a random sample of 20 adults under age 35 is selected, find the probability that exactly 16 have eaten pizza for breakfast.

Source: Harper's Index.

22. Unmarried Women According to survey records, 75.4% of women aged 20–24 have never been married. In a random sample of 250 young women aged 20–24, find the mean, variance, and standard deviation for the number who are or who have been married.

Source: www.infoplease.com

Section 5–4

23. Accuracy Count of Votes After a recent national election, voters were asked how confident they were that votes in their state would be counted accurately. The results are shown below.

46% Very confident	41% Somewhat confident
9% Not very confident	4% Not at all confident

If 10 voters are selected at random, find the probability that 5 would be very confident, 3 somewhat confident, 1 not very confident, and 1 not at all confident.

Source: New York Times.

24. Defective DVDs Before a DVD leaves the factory, it is given a quality control check. The probabilities that a DVD contains 0, 1, or 2 defects are 0.90, 0.06, and 0.04, respectively. In a sample of 12 DVDs, find the probability that 8 have 0 defects, 3 have 1 defect, and 1 has 2 defects.

25. Accounting Errors The probability that an accounting company will make 0, 1, 2, or 3 errors in preparing a yearly budget for a small company is 0.50, 0.28, 0.15, and 0.07 respectively. If 20 companies have the firm prepare their budgets, find the probability that 9 will contain 0 errors, 6 will contain one error, 3 will contain 2 errors, and 2 will contain 3 errors.

26. Lost Luggage in Airlines Transportation officials reported that 8.25 out of every 1000 airline passengers lost luggage during their travels last year. If we randomly select 400 airline passengers, what is the probability that 5 lost some luggage?

Source: U.S. Department of Transportation.

27. Computer Assistance Computer Help Hot Line receives, on average, 6 calls per hour asking for assistance. The distribution is Poisson. For any randomly selected hour, find the probability that the company will receive

a. At least 6 calls
b. 4 or more calls
c. At most 5 calls

28. Boating Accidents The number of boating accidents on Lake Emilie follows a Poisson distribution. The probability of an accident is 0.003. If there are 1000 boats on the lake during a summer month, find the probability that there will be 6 accidents.

29. Drawing Cards If 5 cards are drawn from a deck, find the probability that 2 will be hearts.

30. Car Sales Of the 50 automobiles in a used-car lot, 10 are white. If 5 automobiles are selected to be sold at an auction, find the probability that exactly 2 will be white.

31. Items Donated to a Food Bank At a food bank a case of donated items contains 10 cans of soup, 8 cans of vegetables, and 8 cans of fruit. If 3 cans are selected at random to distribute, find the probability of getting 1 can of vegetables and 2 cans of fruit.

32. Tossing a Die A die is rolled until a 3 is obtained. Find the probability that the first 3 will be obtained on the fourth roll.

33. Selecting a Card A card is selected at random from an ordinary deck and replaced. Find the probability that the first face card will be selected on the fourth draw.

STATISTICS TODAY

Is Pooling Worthwhile?—**Revisited**

In the case of the pooled sample, the probability that only one test will be needed can be determined by using the binomial distribution. The question being asked is, In a sample of 15 individuals, what is the probability that no individual will have the disease? Hence, $n = 15$, $p = 0.05$, and $X = 0$. From Table B in Appendix A, the probability is 0.463, or 46% of the time, only one test will be needed. For screening purposes, then, pooling samples in this case would save considerable time, money, and effort as opposed to testing every individual in the population.

Chapter Quiz

Determine whether each statement is true or false. If the statement is false, explain why.

1. The expected value of a random variable can be thought of as a long-run average.
2. The number of courses a student is taking this semester is an example of a continuous random variable.
3. When the binomial distribution is used, the outcomes must be dependent.
4. A binomial experiment has a fixed number of trials.

Complete these statements with the best answer.

5. Random variable values are determined by ________.

6. The mean for a binomial variable can be found by using the expression ________.

7. One requirement for a probability distribution is that the sum of all the events in the sample space equal ____.

Select the best answer.

8. What is the sum of the probabilities of all outcomes in a probability distribution?

   a. 0 *c.* 1
   b. $\frac{1}{2}$ *d.* It cannot be determined.

9. How many outcomes are there in a binomial experiment?

   a. 0 *c.* 2
   b. 1 *d.* It varies.

10. The number of trials for a binomial experiment

    a. Can be infinite *c.* Is unlimited
    b. Is unchanged *d.* Must be fixed

For exercises 11 through 14, determine if the distribution represents a probability distribution. If not, state why.

11.

X	1	2	3	4	5
$P(X)$	$\frac{1}{7}$	$\frac{2}{7}$	$\frac{2}{7}$	$\frac{3}{7}$	$\frac{2}{7}$

12.

X	3	6	9	12	15
$P(X)$	0.3	0.5	0.1	0.08	0.02

13.

X	50	75	100
$P(X)$	0.5	0.2	0.3

14.

X	4	8	12	16
$P(X)$	$\frac{1}{6}$	$\frac{3}{12}$	$\frac{1}{2}$	$\frac{1}{12}$

15. **Calls for a Fire Company** The number of fire calls the Conestoga Valley Fire Company receives per day is distributed as follows:

Number X	5	6	7	8	9
Probability $P(X)$	0.28	0.32	0.09	0.21	0.10

Construct a graph for the data.

16. **Cell phones per Household** A study was conducted to determine the number of cell phones each household has. The data are shown here.

Number of cell phones	0	1	2	3	4
Frequency	2	30	48	13	7

Construct a probability distribution and draw a graph for the data.

17. **CD Purchases** During a recent CD sale at Matt's Music Store, the number of CDs customers purchased was distributed as follows:

Number X	0	1	2	3	4
Probability $P(X)$	0.10	0.23	0.31	0.27	0.09

Find the mean, variance, and standard deviation of the distribution.

18. **Calls for a Crisis Hot Line** The number of calls received per day at a crisis hot line is distributed as follows:

Number X	30	31	32	33	34
Probability $P(X)$	0.05	0.21	0.38	0.25	0.11

Find the mean, variance, and standard deviation of the distribution.

19. **Selecting a Card** There are 6 playing cards placed face down in a box. They are the 4 of diamonds, the 5 of hearts, the 2 of clubs, the 10 of spades, the 3 of diamonds, and the 7 of hearts. A person selects a card. Find the expected value of the draw.

20. **Selecting a Card** A person selects a card from an ordinary deck of cards. If it is a black card, she wins \$2. If it is a red card between or including 3 and 7, she wins \$10. If it is a red face card, she wins \$25; and if it is a black jack, she wins an extra \$100. Find the expectation of the game.

21. **Carpooling** If 40% of all commuters ride to work in carpools, find the probability that if 8 workers are selected, 5 will ride in carpools.

22. **Employed Women** If 60% of all women are employed outside the home, find the probability that in a sample of 20 women,

    a. Exactly 15 are employed
    b. At least 10 are employed
    c. At most 5 are not employed outside the home

23. **Driver's Exam** If 80% of the applicants are able to pass a driver's proficiency road test, find the mean, variance, and standard deviation of the number of people who pass the test in a sample of 300 applicants.

24. **Meeting Attendance** A history class has 75 members. If there is a 12% absentee rate per class meeting, find the mean, variance, and standard deviation of the number of students who will be absent from each class.

25. **Income Tax Errors** The probability that a person will make 0, 1, 2, or 3 errors on his or her income tax return is 0.50, 0.30, 0.15, and 0.05, respectively. If 30 claims are selected, find the probability that 15 will contain 0 errors, 8 will contain 1 error, 5 will contain 2 errors, and 2 will contain 3 errors.

26. **Quality Control Check** Before a television set leaves the factory, it is given a quality control check. The probability that a television contains 0, 1, or 2 defects is 0.88, 0.08, and 0.04, respectively. In a sample of 16 televisions, find the probability that 9 will have 0 defects, 4 will have 1 defect, and 3 will have 2 defects.

27. Bowling Team Uniforms Among the teams in a bowling league, the probability that the uniforms are all 1 color is 0.45, that 2 colors are used is 0.35, and that 3 or more colors are used is 0.20. If a sample of 12 uniforms is selected, find the probability that 5 contain only 1 color, 4 contain 2 colors, and 3 contain 3 or more colors.

28. Elm Trees If 8% of the population of trees are elm trees, find the probability that in a sample of 100 trees, there are exactly 6 elm trees. Assume the distribution is approximately Poisson.

29. Sports Score Hot Line Calls Sports Scores Hot Line receives, on the average, 8 calls per hour requesting the latest sports scores. The distribution is Poisson in nature. For any randomly selected hour, find the probability that the company will receive

a. At least 8 calls
b. 3 or more calls
c. At most 7 calls

30. Color of Raincoats There are 48 raincoats for sale at a local men's clothing store. Twelve are black. If 6 raincoats are selected to be marked down, find the probability that exactly 3 will be black.

31. Youth Group Officers A youth group has 8 boys and 6 girls. If a slate of 4 officers is selected, find the probability that exactly

a. 3 are girls
b. 2 are girls
c. 4 are boys

32. Blood Types About 4% of the citizens of the United States have type AB blood. If an agency needs type AB blood and donors come in at random, find the probability that the sixth person is the first person with type AB blood.

33. Alcohol Abstainers About 35% of Americans abstain from the consumption of alcohol. If Americans are selected at random, find the probability that the 10th person selected will be the first one who doesn't drink alcohol.

Critical Thinking Challenges

1. **Lottery Numbers** Pennsylvania has a lottery entitled "Big 4." To win, a player must correctly match four digits from a daily lottery in which four digits are selected. Find the probability of winning.
2. **Lottery Numbers** In the Big 4 lottery, for a bet of \$100, the payoff is \$5000. What is the expected value of winning? Is it worth it?
3. **Lottery Numbers** If you played the same four-digit number every day (or any four-digit number for that matter) in the Big 4, how often (in years) would you win, assuming you have average luck?
4. **Chuck-a-Luck** In the game Chuck-a-Luck, three dice are rolled. A player bets a certain amount (say \$1.00) on a number from 1 to 6. If the number appears on 1 die, the person wins \$1.00. If it appears on 2 dice, the person wins \$2.00, and if it appears on all 3 dice, the person wins \$3.00. What are the chances of winning \$1.00? \$2.00? \$3.00?
5. **Chuck-a-Luck** What is the expected value of the game of Chuck-a-Luck if a player bets \$1.00 on one number?

Data Projects

1. **Business and Finance** Assume that a life insurance company would like to make a profit of \$250 on a \$100,000 policy sold to a person whose probability of surviving the year is 0.9985. What premium should the company charge the customer? If the company would like to make a \$250 profit on a \$100,000 policy at a premium of \$500, what is the lowest life expectancy it should accept for a customer?
2. **Sports and Leisure** Baseball, hockey, and basketball all use a seven-game series to determine their championship. Find the probability that with two evenly matched teams a champion will be found in 4 games. Repeat for 5, 6, and 7 games. Look at the historical results for the three sports. How do the actual results compare to the theoretical?
3. **Technology** Use your most recent itemized phone bill for the data in this problem. Assume that incoming and outgoing calls are equal in the population (why is this a reasonable assumption?). This means assume $p = 0.5$. For the number of calls you made last month, what would be the mean number of outgoing calls in a random selection of calls? Also, compute the standard deviation. Was the number of outgoing calls you made an unusual amount given the above? In a selection of 12 calls, what is the probability that less than 3 were outgoing?

4. **Health and Wellness** Use Red Cross data to determine the percentage of the population with an Rh factor that is positive (A+, B+, AB+, or O1 blood types). Use that value for *p*. How many students in your class have a positive Rh factor? Is this an unusual amount?
5. **Politics and Economics** Find out what percentage of citizens in your state is registered to vote. Assuming that this is a binomial variable, what would be the mean number of registered voters in a random group of citizens with a sample size equal to the number of students in your class? Also determine the standard deviation. How many students in your class are registered to vote? Is this an unusual number, given the above?
6. **Your Class** Have each student in class toss 4 coins on her or his desk, and note how many heads are showing. Create a frequency table displaying the results. Compare the frequency table to the theoretical probability distribution for the outcome when 4 coins are tossed. Find the mean for the frequency table. How does it compare with the mean for the probability distribution?

Answers to Applying the Concepts

Section 5–1 Dropping College Courses

1. The random variable under study is the reason for dropping a college course.
2. There were a total of 144 people in the study.
3. The complete table is as follows:

Reason for dropping a college course	Frequency	Percentage
Too difficult	45	31.25
Illness	40	27.78
Change in work schedule	20	13.89
Change of major	14	9.72
Family-related problems	9	6.25
Money	7	4.86
Miscellaneous	6	4.17
No meaningful reason	3	2.08

4. The probability that a student will drop a class because of illness is about 28%. The probability that a student will drop a class because of money is about 5%. The probability that a student will drop a class because of a change of major is about 10%.
5. The information is not itself a probability distribution, but it can be used as one.
6. The categories are not necessarily mutually exclusive, but we treated them as such in computing the probabilities.
7. The categories are not independent.
8. The categories are exhaustive.
9. Since all the probabilities are between 0 and 1, inclusive, and the probabilities sum to 1, the requirements for a discrete probability distribution are met.

Section 5–2 Radiation Exposure

1. The expected value is the mean of a random variable associated with a probability distribution.
2. We would expect variation from the expected value of 3.
3. Answers will vary. One possible answer is that pregnant mothers in that area might be overly concerned upon hearing that the number of cases of kidney problems in newborns was nearly 4 times what was usually expected. Other mothers (particularly those who had taken a statistics course!) might ask for more information about the claim.
4. Answers will vary. One possible answer is that it does seem unlikely to have 11 newborns with kidney problems when we expect only 3 newborns to have kidney problems.
5. The public might be better informed by percentages or rates (e.g. number of newborns with kidney problems per 1000 newborns).
6. The increase of 8 babies born with kidney problems represents a 0.32% increase (less than $\frac{1}{2}$%).
7. Answers will vary. One possible answer is that the percentage increase does not seem to be something to be overly concerned about.

Section 5–3 Unsanitary Restaurants

1. The probability of eating at 3 restaurants with unsanitary conditions out of the 10 restaurants is 0.18792.
2. The probability of eating at 4 or 5 restaurants with unsanitary conditions out of the 10 restaurants is $0.24665 + 0.22199 = 0.46864 \approx 0.469$.
3. To find this probability, you could add the probabilities for eating at 1, 2, . . . , 10 unsanitary restaurants. An easier way to compute the probability is to subtract the probability of eating at no unsanitary restaurants from 1 (using the complement rule).
4. The highest probability for this distribution is 4, but the expected number of unsanitary restaurants that you would eat at is $10 \cdot \frac{3}{7} = 4.29$.
5. The standard deviation for this distribution is $\sqrt{(10)(\frac{3}{7})(\frac{4}{7})} \approx 1.565$.

6. We have two possible outcomes: "success" is eating in an unsanitary restaurant; "failure" is eating in a sanitary restaurant. The probability that one restaurant is unsanitary is independent of the probability that any other restaurant is unsanitary. The probability that a restaurant is unsanitary remains constant at $\frac{3}{7}$. And we are looking at the number of unsanitary restaurants that we eat at out of 10 "trials."

7. The likelihood of success will vary from situation to situation. Just because we have two possible outcomes, this does not mean that each outcome occurs with probability 0.50.

Section 5–4 Rockets and Targets

1. The theoretical values for the number of hits are as follows:

Hits	0	1	2	3	4	5
Regions	227.5	211.3	98.2	30.4	7.1	1.3

2. The actual values are very close to the theoretical values.

3. Since the actual values are close to the theoretical values, it does appear that the rockets were fired at random.

6

The Normal Distribution

© Fuse/Getty Images RF

STATISTICS TODAY

What Is Normal?

Medical researchers have determined so-called normal intervals for a person's blood pressure, cholesterol, triglycerides, and the like. For example, the normal range of systolic blood pressure is 110 to 140. The normal interval for a person's triglycerides is from 30 to 200 milligrams per deciliter (mg/dl). By measuring these variables, a physician can determine if a patient's vital statistics are within the normal interval or if some type of treatment is needed to correct a condition and avoid future illnesses. The question then is, How does one determine the so-called normal intervals? See Statistics Today—Revisited at the end of the chapter.

In this chapter, you will learn how researchers determine normal intervals for specific medical tests by using a normal distribution. You will see how the same methods are used to determine the lifetimes of batteries, the strength of ropes, and many other traits.

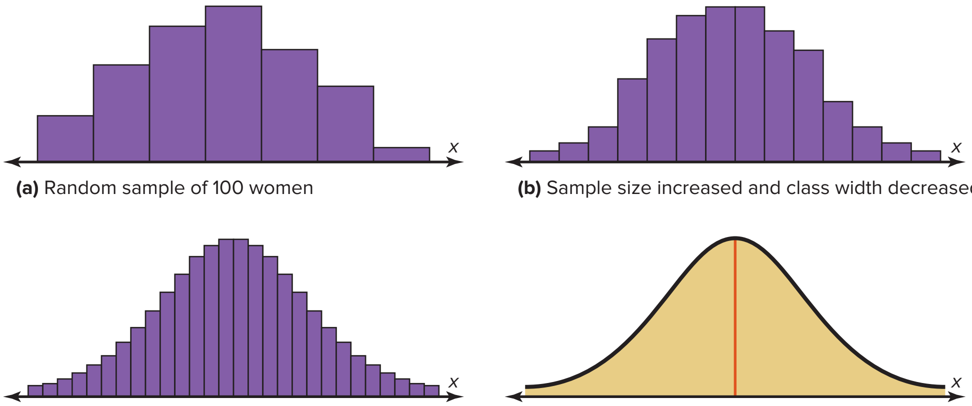
OUTLINE

OBJECTIVES

After completing this chapter, you should be able to:

1. Identify the properties of a normal distribution.
2. Identify distributions as symmetric or skewed.
3. Find the area under the standard normal distribution, given various z values.
4. Find probabilities for a normally distributed variable by transforming it into a standard normal variable.
5. Find specific data values for given percentages, using the standard normal distribution.
6. Use the central limit theorem to solve problems involving sample means for large samples.
7. Use the normal approximation to compute probabilities for a binomial variable.

FIGURE 6–1
Histograms and Normal Model for the Distribution of Heights of Adult Women



(a) Random sample of 100 women
(b) Sample size increased and class width decreased
(c) Sample size increased and class width decreased further
(d) Normal distribution for the population

Introduction

Historical Note
The name *normal curve* was used by several statisticians, namely, Francis Galton, Charles Sanders, Wilhelm Lexis, and Karl Pearson near the end of the 19th century.

Random variables can be either discrete or continuous. Discrete variables and their distributions were explained in Chapter 5. Recall that a discrete variable cannot assume all values between any two given values of the variables. On the other hand, a continuous variable can assume all values between any two given values of the variables. Examples of continuous variables are the height of adult men, body temperature of rats, and cholesterol level of adults. Many continuous variables, such as the examples just mentioned, have distributions that are bell-shaped, and these are called *approximately normally distributed variables.* For example, if a researcher selects a random sample of 100 adult women, measures their heights, and constructs a histogram, the researcher gets a graph similar to the one shown in Figure 6–1(a). Now, if the researcher increases the sample size and decreases the width of the classes, the histograms will look like the ones shown in Figure 6–1(b) and (c). Finally, if it were possible to measure exactly the heights of all adult females in the United States and plot them, the histogram would approach what is called a *normal distribution curve,* as shown in Figure 6–1(d). This distribution is also known as a *bell curve* or a *Gaussian distribution curve,* named for the German mathematician Carl Friedrich Gauss (1777–1855), who derived its equation.

No variable fits a normal distribution perfectly, since a normal distribution is a theoretical distribution. However, a normal distribution can be used to describe many variables, because the deviations from a normal distribution are very small. This concept will be explained further in Section 6–1.

This chapter will also present the properties of a normal distribution and discuss its applications. Then a very important fact about a normal distribution called the *central limit theorem* will be explained. Finally, the chapter will explain how a normal distribution curve can be used as an approximation to other distributions, such as the binomial distribution. Since a binomial distribution is a discrete distribution, a correction for continuity may be employed when a normal distribution is used for its approximation.

6–1 Normal Distributions

In mathematics, curves can be represented by equations. For example, the equation of the circle shown in Figure 6–2 is $x^2 + y^2 = r^2$, where r is the radius. A circle can be used to represent many physical objects, such as a wheel or a gear. Even though it is not possible

to manufacture a wheel that is perfectly round, the equation and the properties of a circle can be used to study many aspects of the wheel, such as area, velocity, and acceleration. In a similar manner, the theoretical curve, called a *normal distribution curve,* can be used to study many variables that are not perfectly normally distributed but are nevertheless approximately normal.

If a random variable has a probability distribution whose graph is continuous, bell-shaped, and symmetric, it is called a **normal distribution**. The graph is called a *normal distribution* curve.

The mathematical equation for a normal distribution is

$$y = \frac{e^{-(X-\mu)^2/(2\sigma^2)}}{\sigma\sqrt{2\pi}}$$

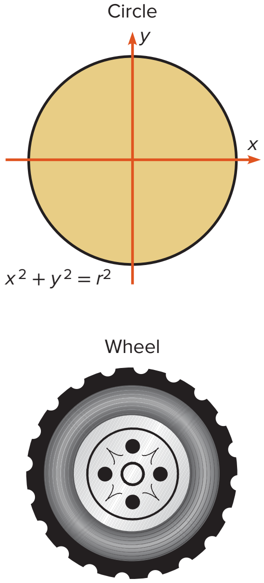
where $e \approx 2.718$ ($\approx$ means "is approximately equal to")

$\pi \approx 3.14$

$\mu =$ population mean

$\sigma =$ population standard deviation

FIGURE 6–2
Graph of a Circle and an Application



This equation may look formidable, but in applied statistics, tables or technology is used for specific problems instead of the equation.

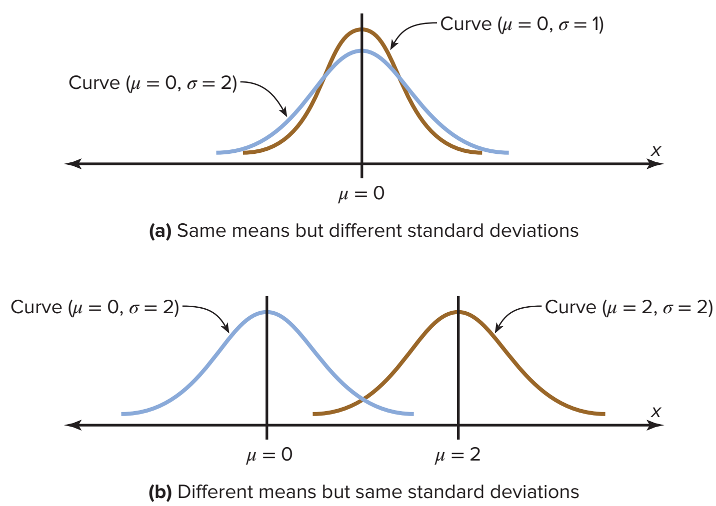
Another important consideration in applied statistics is that the area under a normal distribution curve is used more often than the values on the *y* axis. Therefore, when a normal distribution is pictured, the *y* axis is sometimes omitted.

Circles can be different sizes, depending on their diameters (or radii), and can be used to represent wheels of different sizes. Likewise, normal curves have different shapes and can be used to represent different variables.

The shape and position of a normal distribution curve depend on two parameters, the *mean* and the *standard deviation.* Each normally distributed variable has its own normal distribution curve, which depends on the values of the variable's mean and standard deviation.

Suppose one normally distributed variable has $\mu = 0$ and $\sigma = 1$, and another normally distributed variable has $\mu = 0$ and $\sigma = 2$. As you can see in Figure 6–3(a), when the value

FIGURE 6–3
Shapes of Normal Distributions



OBJECTIVE 1

Identify the properties of a normal distribution.

of the standard deviation increases, the shape of the curve spreads out. If one normally distributed variable has $\mu = 0$ and $\sigma = 2$ and another normally distributed variable has $\mu = 2$, and $\sigma = 2$, then the shapes of the curve are the same, but the curve with $\mu = 2$ moves 2 units to the right. See Figure 6–3(b).

The properties of a normal distribution, including those mentioned in the definition, are explained next.

Historical Notes

The discovery of the equation for a normal distribution can be traced to three mathematicians. In 1733, the French mathematician Abraham DeMoivre derived an equation for a normal distribution based on the random variation of the number of heads appearing when a large number of coins were tossed. Not realizing any connection with the naturally occurring variables, he showed this formula to only a few friends. About 100 years later, two mathematicians, Pierre Laplace in France and Carl Gauss in Germany, derived the equation of the normal curve independently and without any knowledge of DeMoivre's work. In 1924, Karl Pearson found that DeMoivre had discovered the formula before Laplace or Gauss.

Summary of the Properties of the Theoretical Normal Distribution

1. A normal distribution curve is bell-shaped.
2. The mean, median, and mode are equal and are located at the center of the distribution.
3. A normal distribution curve is unimodal (i.e., it has only one mode).
4. The curve is symmetric about the mean, which is equivalent to saying that its shape is the same on both sides of a vertical line passing through the center.
5. The curve is continuous; that is, there are no gaps or holes. For each value of X, there is a corresponding value of Y.
6. The curve never touches the x axis. Theoretically, no matter how far in either direction the curve extends, it never meets the x axis—but it gets increasingly close.
7. The total area under a normal distribution curve is equal to 1.00, or 100%. This fact may seem unusual, since the curve never touches the x axis, but one can prove it mathematically by using calculus. (The proof is beyond the scope of this text.)
8. The area under the part of a normal curve that lies within 1 standard deviation of the mean is approximately 0.68, or 68%; within 2 standard deviations, about 0.95, or 95%; and within 3 standard deviations, about 0.997, or 99.7%. See Figure 6–4, which also shows the area in each region.

The values given in item 8 of the summary follow the *empirical rule* for data given in Section 3–2.

You must know these properties in order to solve problems involving distributions that are approximately normal.

Recall from Chapter 2 that the graphs of distributions can have many shapes. When the data values are evenly distributed about the mean, a distribution is said to be a **symmetric distribution.** (A normal distribution is symmetric.) Figure 6–5(a) shows a symmetric distribution. When the majority of the data values fall to the left or right of the mean, the distribution is said to be *skewed.* When the majority of the data values fall to the

FIGURE 6–4
Areas Under a Normal Distribution Curve

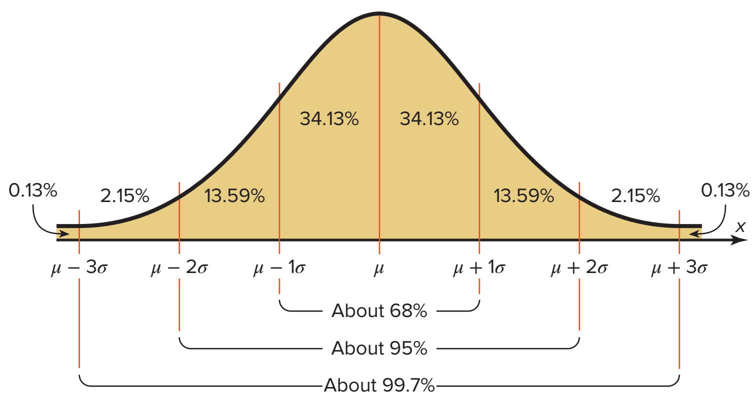
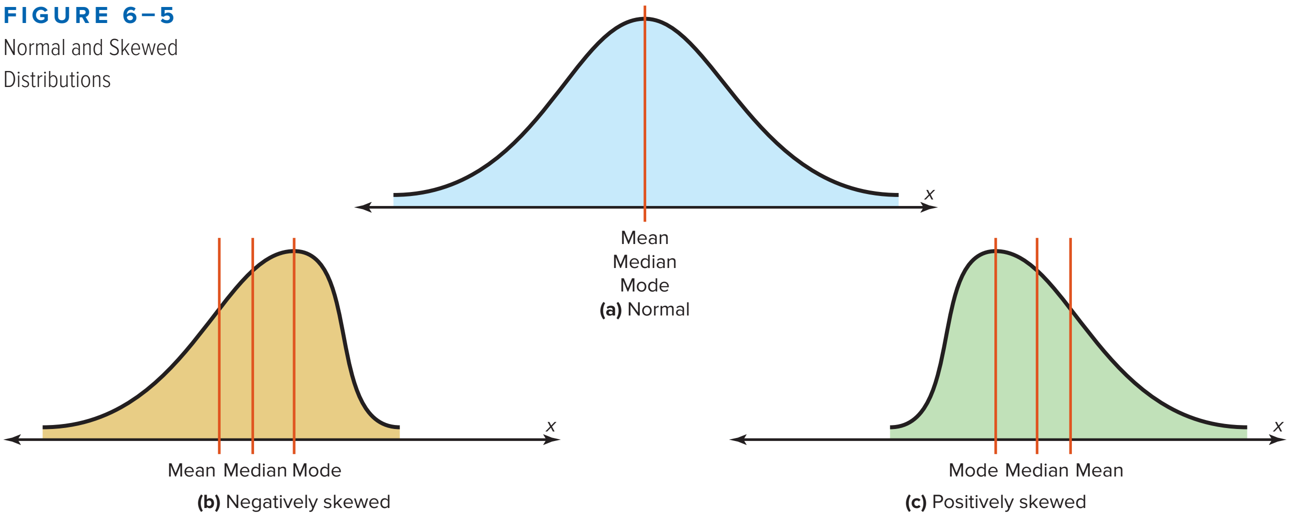


FIGURE 6–5
Normal and Skewed Distributions



OBJECTIVE 2
Identify distributions as symmetric or skewed.

right of the mean, the distribution is said to be a **negatively or left-skewed distribution.** The mean is to the left of the median, and the mean and the median are to the left of the mode. See Figure 6–5(b). When the majority of the data values fall to the left of the mean, a distribution is said to be a **positively or right-skewed distribution.** The mean falls to the right of the median, and both the mean and the median fall to the right of the mode. See Figure 6–5(c).

The "tail" of the curve indicates the direction of skewness (right is positive, left is negative). These distributions can be compared with the ones shown in Figure 3–1. Both types follow the same principles.

The Standard Normal Distribution

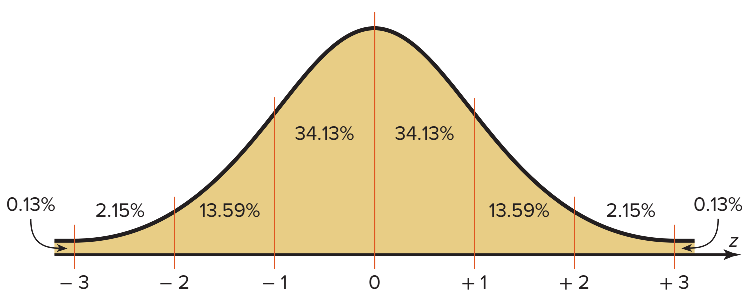
Since each normally distributed variable has its own mean and standard deviation, as stated earlier, the shape and location of these curves will vary. In practical applications, then, you would have to have a table of areas under the curve for each variable. To simplify this situation, statisticians use what is called the *standard normal distribution.*

OBJECTIVE 3
Find the area under the standard normal distribution, given various z values.

The **standard normal distribution** is a normal distribution with a mean of 0 and a standard deviation of 1.

The standard normal distribution is shown in Figure 6–6.

FIGURE 6–6
Standard Normal Distribution



The values under the curve indicate the proportion of area in each section. For example, the area between the mean and 1 standard deviation above or below the mean is about 0.3413, or 34.13%.

The formula for the standard normal distribution is

$$y = \frac{e^{-z^2/2}}{\sqrt{2\pi}}$$

All normally distributed variables can be transformed into the standard normally distributed variable by using the formula for the standard score:

$$z = \frac{\text{value} - \text{mean}}{\text{standard deviation}} \qquad \text{or} \qquad z = \frac{X - \mu}{\sigma}$$

This is the same formula used in Section 3–3. The use of this formula will be explained in Section 6–3.

As stated earlier, the area under a normal distribution curve is used to solve practical application problems, such as finding the percentage of adult women whose height is between 5 feet 4 inches and 5 feet 7 inches, or finding the probability that a new battery will last longer than 4 years. Hence, the major emphasis of this section will be to show the procedure for finding the area under the standard normal distribution curve for any z value. The applications will be shown in Section 6–2. Once the X values are transformed by using the preceding formula, they are called z values. The ***z* value** or ***z* score** is actually the number of standard deviations that a particular X value is away from the mean. Table E in Appendix A gives the area (to four decimal places) under the standard normal curve for any z value from −3.49 to 3.49.

Interesting Fact

Bell-shaped distributions occurred quite often in early coin-tossing and die-rolling experiments.

Finding Areas Under the Standard Normal Distribution Curve

For the solution of problems using the standard normal distribution, a two-step process is recommended with the use of the Procedure Table shown.

The two steps are as follows:

Step 1 Draw the normal distribution curve and shade the area.

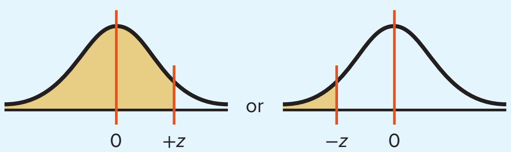
Step 2 Find the appropriate figure in the Procedure Table and follow the directions given.

There are three basic types of problems, and all three are summarized in the Procedure Table. Note that this table is presented as an aid in understanding how to use the standard normal distribution table and in visualizing the problems. After learning the procedures, you should not find it necessary to refer to the Procedure Table for every problem.

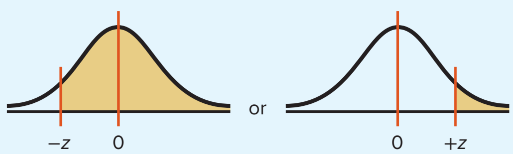
Procedure Table

Finding the Area Under the Standard Normal Distribution Curve

1. To the left of any z value:
   Look up the z value in the table and use the area given.



2. To the right of any z value:
   Look up the z value and subtract the area from 1.



3. Between any two z values:
   Look up both z values and subtract the corresponding areas.

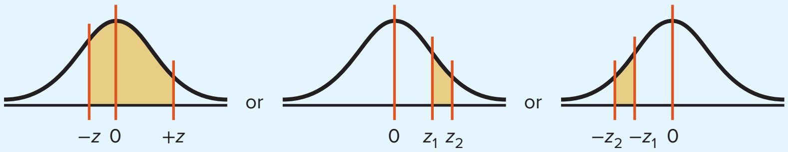


FIGURE 6–7
Table E Area Value for $z = 1.39$

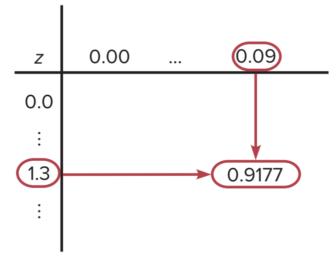


Table E in Appendix A gives the area under the normal distribution curve to the left of any z value given in two decimal places. For example, the area to the left of a z value of 1.39 is found by looking up 1.3 in the left column and 0.09 in the top row. Where the row and column lines meet gives an area of 0.9177. See Figure 6–7.

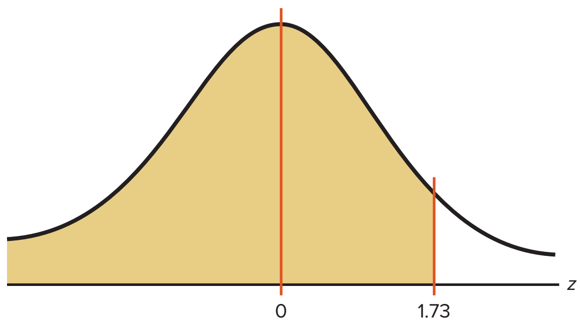
EXAMPLE 6–1

Find the area under the standard normal distribution curve to the left of $z = 1.73$.

SOLUTION

Step 1 Draw the figure. The desired area is shown in Figure 6–8.

FIGURE 6–8
Area Under the Standard Normal Distribution Curve for Example 6–1



Step 2 We are looking for the area under the standard deviation distribution curve to the left of $z = 1.73$. Since this is an example of the first case, look up the area in the table. It is 0.9582. Hence, 95.82% of the area is to the left of $z = 1.73$.

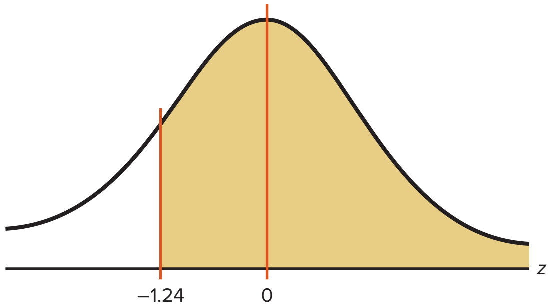
EXAMPLE 6–2

Find the area under the standard normal distribution curve to the right of $z = -1.24$.

SOLUTION

Step 1 Draw the figure. The desired area is shown in Figure 6–9.

FIGURE 6–9
Area Under the Standard Normal Distribution Curve for Example 6–2



Step 2 We are looking for the area to the right of $z = -1.24$. This is an example of the second case. Look up the area for $z = -1.24$. It is 0.1075. Subtract the area from 1.0000: $1.0000 - 0.1075 = 0.8925$. Hence, 89.25% of the area under the standard normal distribution curve is to the right of -1.24.

Note: In this situation, we subtract the area 0.1075 from1.0000 because the table gives the area up to -1.24. So to get the area under the curve to the right of -1.24, subtract the area 0.1075 from 1.0000, the total area under the standard normal distribution curve.

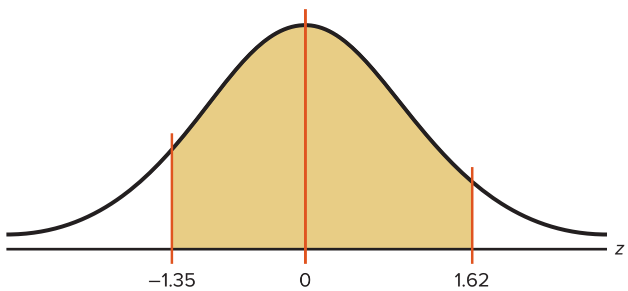
EXAMPLE 6–3

Find the area under the standard normal distribution curve between $z = 1.62$ and $z = -1.35$.

SOLUTION

Step 1 Draw the figure as shown. The desired area is shown in Figure 6–10.

FIGURE 6–10
Area Under the Standard Normal Distribution Curve for Example 6–3



Step 2 Since the area desired is between two given z values, look up the areas corresponding to the two z values and subtract the smaller area from the larger area. (Do not subtract the z values.) The area for $z = 1.62$ is 0.9474, and the area for $z = -1.35$ is 0.0885. The area between the two z values is $0.9474 - 0.0885 = 0.8589$, or 85.89%.

A Normal Distribution Curve as a Probability Distribution Curve

A normal distribution curve can be used as a probability distribution curve for normally distributed variables. Recall that a normal distribution is a *continuous distribution,* as opposed to a discrete probability distribution, as explained in Chapter 5. The fact that it is continuous means that there are no gaps in the curve. In other words, for every z value on the x axis, there is a corresponding height, or frequency, value.

The area under the standard normal distribution curve can also be thought of as a probability or as the proportion of the population with a given characteristic. That is, if it were possible to select a z value at random, the probability of choosing one, say, between 0 and 2.00 would be the same as the area under the curve between 0 and 2.00. In this case, the area is 0.4772. Therefore, the probability of randomly selecting a z value between 0 and 2.00 is 0.4772. The problems involving probability are solved in the same manner as the previous examples involving areas in this section. For example, if the problem is to find the probability of selecting a z value between 2.25 and 2.94, solve it by using the method shown in case 3 of the Procedure Table.

For probabilities, a special notation is used to denote the probability of a standard normal variable z. For example, if the problem is to find the probability of any z value between 0 and 2.32, this probability is written as $P(0 < z < 2.32)$.

Note: In a continuous distribution, the probability of any exact z value is 0 since the area would be represented by a vertical line above the value. But vertical lines in theory have no area. So $P(a \leq z \leq b) = P(a < z < b)$.

EXAMPLE 6–4

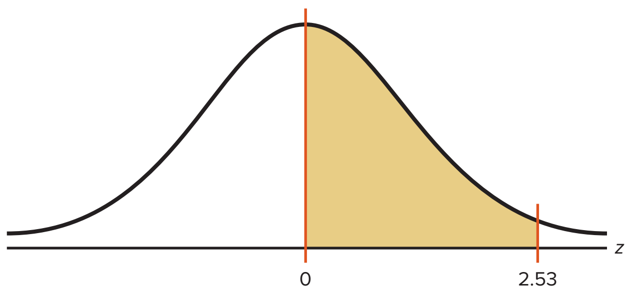
Find the probability for each. (Assume this is a standard normal distribution.)

a. $P(0 < z < 2.53)$ *b.* $P(z < 1.73)$ *c.* $P(z > 1.98)$

SOLUTION

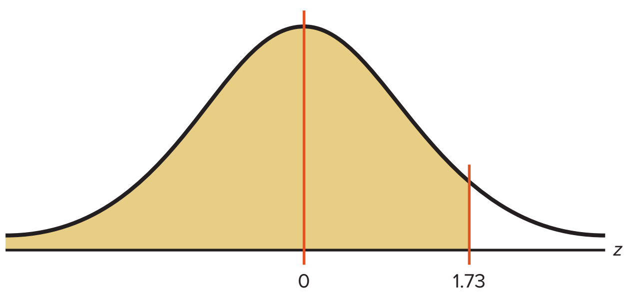
a. $P(0 < z < 2.53)$ is used to find the area under the standard normal distribution curve between $z = 0$ and $z = 2.53$. First, draw the curve and shade the desired area. This is shown in Figure 6–11. Second, find the area in Table E corresponding to $z = 2.53$. It is 0.9943. Third, find the area in Table E corresponding to $z = 0$. It is 0.5000. Finally, subtract the two areas: $0.9943 - 0.5000 = 0.4943$. Hence, the probability is 0.4943, or 49.43%.

FIGURE 6–11
Area Under the Standard Normal Distribution Curve for Part *a* of Example 6–4



b. $P(z < 1.73)$ is used to find the area under the standard normal distribution curve to the left of $z = 1.73$. First, draw the curve and shade the desired area. This is shown in Figure 6–12. Second, find the area in Table E corresponding to 1.73. It is 0.9582. Hence, the probability of obtaining a z value less than 1.73 is 0.9582, or 95.82%.

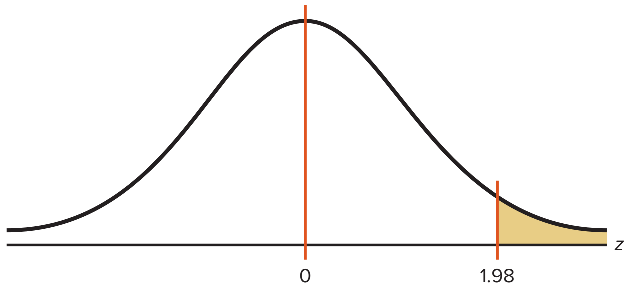
FIGURE 6–12
Area Under the Standard Normal Distribution Curve for Part *b* of Example 6–4



c. $P(z > 1.98)$ is used to find the area under the standard normal distribution curve to the right of $z = 1.98$. First, draw the curve and shade the desired area.

See Figure 6–13. Second, find the area corresponding to $z = 1.98$ in Table E. It is 0.9761. Finally, subtract this area from 1.0000. It is $1.0000 - 0.9761 = 0.0239$. Hence, the probability of obtaining a z value greater than 1.98 is 0.0239, or 2.39%.

FIGURE 6–13
Area Under the Standard Normal Distribution Curve for Part *c* of Example 6–4



Sometimes, one must find a specific z value for a given area under the standard normal distribution curve. The procedure is to work backward, using Table E.

Since Table E is cumulative, it is necessary to locate the cumulative area up to a given z value. Example 6–5 shows this.

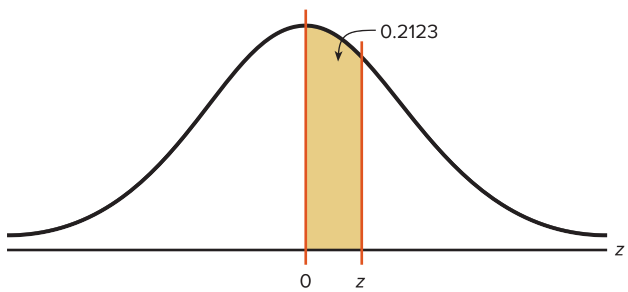
EXAMPLE 6–5

Find the z value such that the area under the standard normal distribution curve between 0 and the z value is 0.2123.

SOLUTION

Draw the figure. The area is shown in Figure 6–14.

FIGURE 6–14
Area Under the Standard Normal Distribution Curve for Example 6–5



In this case it is necessary to add 0.5000 to the given area of 0.2123 to get the cumulative area of 0.7123. Look up the area in Table E. The value in the left column is 0.5, and the top value is 0.06. Add these two values to get $z = 0.56$. See Figure 6–15.

FIGURE 6–15
Finding the z Value from Table E for Example 6–5

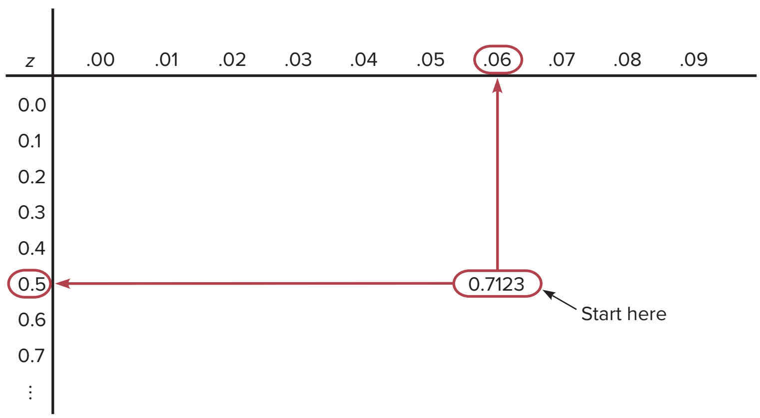
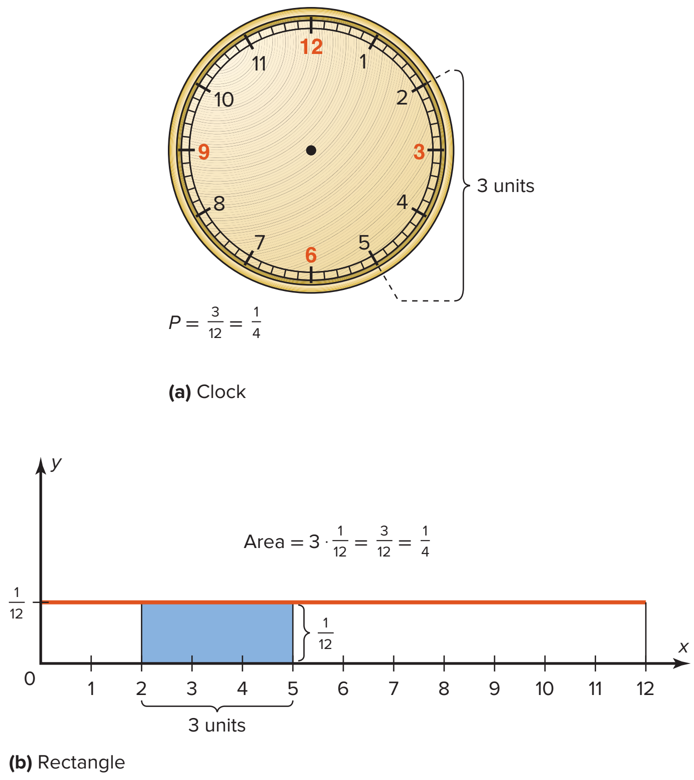


FIGURE 6–16
The Relationship Between Area and Probability



If the exact area cannot be found, use the closest value. For example, if you wanted to find the z value for an area 0.9241, the closest area is 0.9236, which gives a z value of 1.43. See Table E in Appendix C.

The rationale for using an area under a continuous curve to determine a probability can be understood by considering the example of a watch that is powered by a battery. When the battery goes dead, what is the probability that the minute hand will stop somewhere between the numbers 2 and 5 on the face of the watch? In this case, the values of the variable constitute a continuous variable since the hour hand can stop anywhere on the dial's face between 0 and 12 (one revolution of the minute hand). Hence, the sample space can be considered to be 12 units long, and the distance between the numbers 2 and 5 is $5 - 2$, or 3 units. Hence, the probability that the minute hand stops on a number between 2 and 5 is $\frac{3}{12} = \frac{1}{4}$. See Figure 6–16(a).

The problem could also be solved by using a graph of a continuous variable. Let us assume that since the watch can stop anytime at random, the values where the minute hand would land are spread evenly over the range of 0 through 12. The graph would then consist of a *continuous uniform distribution* with a range of 12 units. Now if we required the area under the curve to be 1 (like the area under the standard normal distribution), the height of the rectangle formed by the curve and the x axis would need to be $\frac{1}{12}$. The reason is that the area of a rectangle is equal to the base times the height. If the base is 12 units long, then the height has to be $\frac{1}{12}$ since $12 \cdot \frac{1}{12} = 1$.

The area of the rectangle with a base from 2 through 5 would be $3 \cdot \frac{1}{12}$, or $\frac{1}{4}$. See Figure 6–16(b). Notice that the area of the small rectangle is the same as the probability found previously. Hence, the area of this rectangle corresponds to the probability of this event. The same reasoning can be applied to the standard normal distribution curve shown in Example 6–5.

Finding the area under the standard normal distribution curve is the first step in solving a wide variety of practical applications in which the variables are normally distributed. Some of these applications will be presented in Section 6–2.

Applying the Concepts 6–1

Assessing Normality

Many times in statistics it is necessary to see if a set of data values is approximately normally distributed. There are special techniques that can be used. One technique is to draw a histogram for the data and see if it is approximately bell-shaped. (*Note:* It does not have to be exactly symmetric to be bell-shaped.)

The numbers of branches of the 50 top banks are shown.

67	84	80	77	97	59	62	37	33	42
36	54	18	12	19	33	49	24	25	22
24	29	9	21	21	24	31	17	15	21
13	19	19	22	22	30	41	22	18	20
26	33	14	14	16	22	26	10	16	24

1. Construct a frequency distribution for the data.
2. Construct a histogram for the data.
3. Describe the shape of the histogram.
4. Based on your answer to question 3, do you feel that the distribution is approximately normal?

In addition to the histogram, distributions that are approximately normal have about 68% of the values fall within 1 standard deviation of the mean, about 95% of the data values fall within 2 standard deviations of the mean, and almost 100% of the data values fall within 3 standard deviations of the mean. (See Figure 6–5.)

5. Find the mean and standard deviation for the data.
6. What percent of the data values fall within 1 standard deviation of the mean?
7. What percent of the data values fall within 2 standard deviations of the mean?
8. What percent of the data values fall within 3 standard deviations of the mean?
9. How do your answers to questions 6, 7, and 8 compare to 68, 95, and 100%, respectively?
10. Does your answer help support the conclusion you reached in question 4? Explain.

(More techniques for assessing normality are explained in Section 6–2.)
See pages 367–368 for the answers.

Exercises 6–1

1. What are the characteristics of a normal distribution?

2. Why is the standard normal distribution important in statistical analysis?

3. What is the total area under the standard normal distribution curve?

4. What percentage of the area falls below the mean? Above the mean?

5. About what percentage of the area under the normal distribution curve falls within 1 standard deviation above and below the mean? 2 standard deviations? 3 standard deviations?

6. What are two other names for a normal distribution?

For Exercises 7 through 26, find the area under the standard normal distribution curve.

7. Between $z = 0$ and $z = 1.07$

8. Between $z = 0$ and $z = 1.77$

9. Between $z = 0$ and $z = 1.93$

10. Between $z = 0$ and $z = -0.32$

11. To the right of $z = 0.37$

12. To the right of $z = 2.01$

13. To the left of $z = -1.87$

14. To the left of $z = -0.75$

15. Between $z = 1.09$ and $z = 1.83$

16. Between $z = 1.23$ and $z = 1.90$

17. Between $z = -1.46$ and $z = -1.77$

18. Between $z = -0.96$ and $z = -0.36$

19. Between $z = -1.46$ and $z = -1.98$

20. Between $z = 0.24$ and $z = -1.12$

21. To the left of $z = 1.12$

22. To the left of $z = 1.31$

23. To the right of $z = -0.18$

24. To the right of $z = -1.92$

25. To the right of $z = 1.92$ and to the left of $z = -0.44$

26. To the left of $z = -2.15$ and to the right of $z = 1.62$

In Exercises 27 through 40, find the probabilities for each, using the standard normal distribution.

27. $P(0 < z < 0.95)$

28. $P(0 < z < 1.96)$

29. $P(-1.38 < z < 0)$

30. $P(-1.23 < z < 0)$

31. $P(z > 2.33)$

32. $P(z > 0.82)$

33. $P(z < -1.51)$

34. $P(z < -1.77)$

35. $P(-2.07 < z < 1.88)$

36. $P(-0.20 < z < 1.56)$

37. $P(1.56 < z < 2.13)$

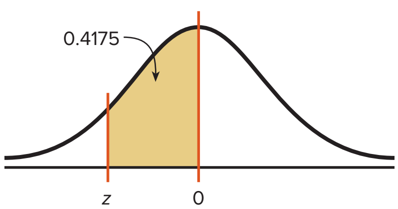
38. $P(1.12 < z < 1.43)$

39. $P(z < 1.42)$

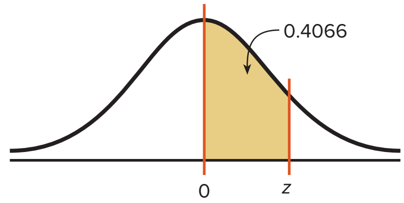
40. $P(z > -1.43)$

For Exercises 41 through 46, find the z value that corresponds to the given area.

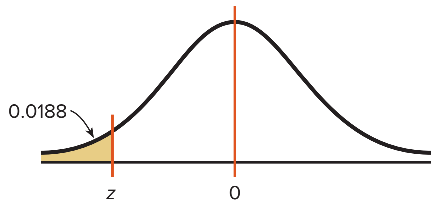
41.



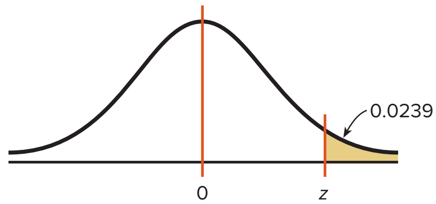
42.



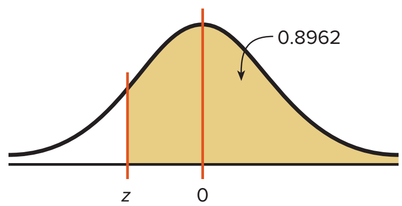
43.



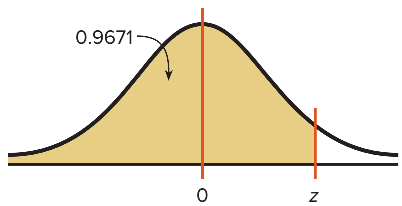
44.



45.



46.



47. Find the z value to the left of the mean so that

a. 98.87% of the area under the distribution curve lies to the right of it.

b. 82.12% of the area under the distribution curve lies to the right of it.

c. 60.64% of the area under the distribution curve lies to the right of it.

48. Find the z value to the right of the mean so that

a. 54.78% of the area under the distribution curve lies to the left of it.

b. 69.85% of the area under the distribution curve lies to the left of it.

c. 88.10% of the area under the distribution curve lies to the left of it.

49. Find two z values, one positive and one negative, that are equidistant from the mean so that the areas in the two tails add to the following values.

a. 5%

b. 10%

c. 1%

50. Find two z values so that 48% of the middle area is bounded by them.

Extending the Concepts

51. Find $P(-1 < z < 1)$, $P(-2 < z < 2)$, and $P(-3 < z < 3)$. How do these values compare with the empirical rule?

52. In the standard normal distribution, find the values of z for the 75th, 80th, and 92nd percentiles.

For Exercises 53–56, z_0 is the statistical notation for an unknown z value. It serves that same function as x does in an algebraic equation.

53. Find z_0 such that $P(-1.2 < z < z_0) = 0.8671$.

54. Find z_0 such that $P(z_0 < z < 2.5) = 0.7672$.

55. Find z_0 such that the area between z_0 and $z = -0.5$ is 0.2345 (two answers).

56. Find z_0 such that $P(-z_0 < z < z_0) = 0.76$.

57. Find the equation for the standard normal distribution by substituting 0 for μ and 1 for σ in the equation

$$y = \frac{e^{-(X-\mu)^2/(2\sigma^2)}}{\sigma\sqrt{2\pi}}$$

58. Graph by hand the standard normal distribution by using the formula derived in Exercise 57. Let $\pi \approx 3.14$ and $e \approx 2.718$. Use X values of $-2, -1.5, -1, -0.5, 0, 0.5, 1, 1.5$, and 2. (Use a calculator to compute the y values.)

59. Find $P(z < 2.3 \text{ or } z > -1.2)$.

60. Find $P(z > 2.3 \text{ and } z < -1.2)$.

Technology Step by Step

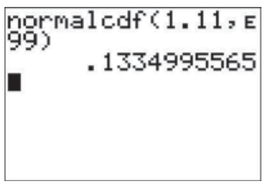
TI–84 Plus

Step by Step

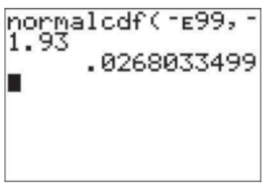
Standard Normal Random Variables

To find the probability for a standard normal random variable:
Press **2nd [DISTR],** then **2** for normalcdf(
The form is normalcdf(lower z score, upper z score).
Use E99 for ∞ (infinity) and $-$E99 for $-\infty$ (negative infinity). Press **2nd [EE]** to get E.

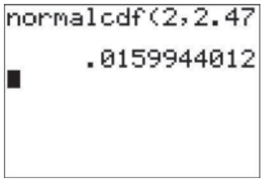
Example: Area to the right of $z = 1.11$
normalcdf(1.11,E99)



Example: Area to the left of $z = -1.93$
normalcdf(−E99,−1.93)



Example: Area between $z = 2.00$ and $z = 2.47$
normalcdf(2.00,2.47)



To find the percentile for a standard normal random variable:
Press **2nd [DISTR],** then **3** for the invNorm(
The form is invNorm(area to the left of z score)

Example: Find the z score such that the area under the standard normal curve to the left of it is 0.7123.
invNorm(.7123)

EXCEL
Step by Step

The Standard Normal Distribution

Finding Areas under the Standard Normal Distribution Curve

Example XL6–1

Find the area to the left of $z = 1.99$.
In a blank cell type: =NORMSDIST(1.99)
Answer: 0.976705

Example XL6–2

Find the area to the right of $z = -2.04$.
In a blank cell type: = 1-NORMSDIST(−2.04)
Answer: 0.979325

Example XL6–3

Find the area between $z = -2.04$ and $z = 1.99$.
In a blank cell type: =NORMSDIST(1.99) − NORMSDIST(−2.04)
Answer: 0.956029

Finding a *z* Value Given an Area Under the Standard Normal Distribution Curve

Example XL6–4

Find a z score given the cumulative area (area to the left of z) is 0.0250.
In a blank cell type: =NORMSINV(.025)
Answer: −1.95996

Example XL6–5

Find a z score, given the area to the right of z is 0.4567.
We must find the z score corresponding to a cumulative area 1 − 0.4567.
In a blank cell type: =NORMSINV(1 − .4567)
Answer: 0.108751

Example XL6–6

Find a (positive) z score given the area between $-z$ and z is 0.98.
We must find the z score corresponding to a cumulative area 0.98 + 0.01 or 0.99
In a blank cell type: =NORMSINV(.99)
Answer: 2.326348

Note: We could also type: = −NORMSINV(.01) to find the z score for Example XL6–6.

MINITAB
Step by Step

Normal Distributions with MINITAB

Finding the area above an interval for a normal distribution is synonymous with finding a probability. We never find the probability that X or Z is exactly equal to a value. We always find the probability that the value is some interval, some range of values. These MINITAB instructions will show you how to get the same results that you would by using Table E, Standard Normal Distribution. Five cases follow.

Open MINITAB. It doesn't matter if something is in the worksheet or not.

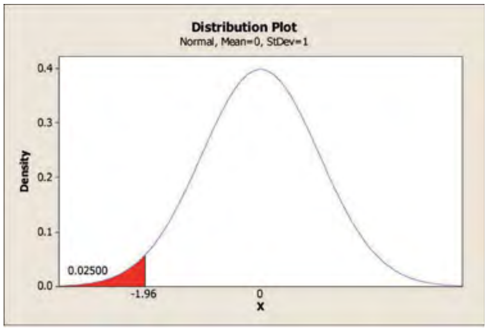
Case 1: Find the Probability That *z* Is to the Left of a Given Value

Find the area to the left of $z = -1.96$. This is the same as saying $z < -1.96$, represented by the number line to the left of −1.96.

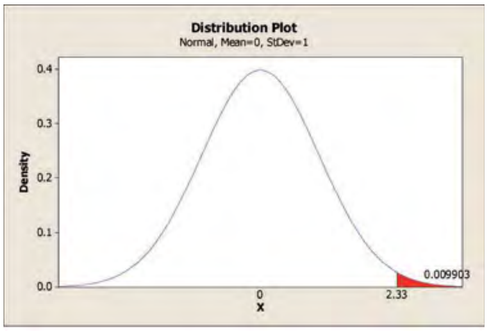
1. Select **Graph>Probability Distribution Plot>View Probability,** then click [OK].
   a) The distribution should be Normal with the Mean set to **0.0** and the Standard deviation set to **1.0.**
   b) Choose the button for Shaded Area, then select the ratio button for X Value.
   c) Click the picture for Left Tail.
   d) Type in the Z value of −1.96 and click [OK].

$$P(Z < -1.96) = 0.02500.$$

You can copy and paste the graph into a document if you like. To copy, right-click on the gray area of the graph in MINITAB. Options will be given for copying the graph.



Case 1 $Z < -1.96$



Case 2 $Z > 2.33$

Case 2: Find the Probability That *z* Is to the Right of a Given Value

Use these instructions for positive or negative values of z.
Find the area **to the right of $z = +2.33$.** This is the same as $Z > 2.33$.

2. Click the icon for **Edit Last Dialog** box or select **Graph>Probability Distribution Plot>View Probability** and click [OK].
   a) The distribution should be normal with the mean set to 0.0 and the standard deviation set to 1.0.

b) Choose the tab for Shaded Area, then select the ratio button for *X* Value.

c) Click the picture for Right Tail.

d) Type in the *Z* value of 2.33 and click [OK].

$$P(X > +2.33) = 0.009903.$$

Case 3: Find the Probability That *Z* Is between Two Values

Find the area if *z* is between −1.11 and +0.24.

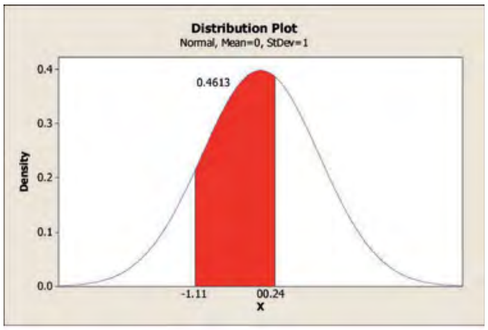
3. Click the icon for **Edit Last Dialog** box or select **Graph>Probability Distribution Plot>View Probability** and click [OK].

a) The distribution should be Normal with the Mean set to **0.0** and the Standard deviation set to **1.0.**

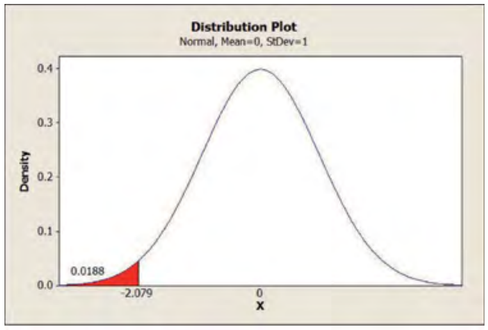
b) Choose the tab for Shaded Area, then *X* Value.

c) Click the picture for Middle.

d) Type in the smaller value −1.11 for *X* value 1 and then the larger value 0.24 for the *X* value 2. Click [OK]. $P(-1.11 < Z < +0.24) = 0.4613$. Remember that smaller values are to the left on the number line.



Case 3



Case 4

Case 4: Find *z* if the Area Is Given

If the area to the left of some *z* value is 0.0188, find the *z* value.

4. Select **Graph>Probability Distribution Plot>View Probability** and click [OK].

a) The distribution should be Normal with the Mean set to 0.0 and the Standard deviation set to 1.0.

b) Choose the tab for Shaded Area and then the ratio button for Probability.

c) Select Left Tail.

d) Type in 0.0188 for probability and then click [OK]. The *z* value is −2.079.

$$P(Z < -2.079) = 0.0188.$$

Case 5: Find Two *z* Values, One Positive and One Negative (Same Absolute Value), so That the Area in the Middle is 0.95

5. Select **Graph>Probability Distribution Plot>View Probability** or click the **Edit Last Dialog** icon.

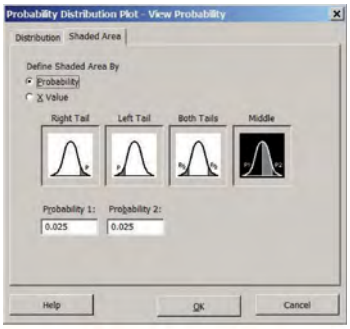
a) The distribution should be Normal with the Mean set to **0.0** and the Standard deviation set to **1.0.**

b) Choose the tab for Shaded Area, then select the ratio button for Probability.

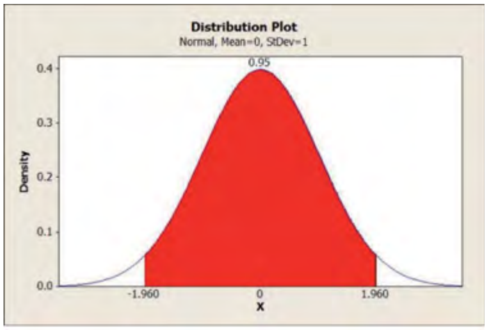
c) Select Middle. You will need to know the area in each tail of the distribution. Subtract 0.95 from 1, then divide by 2. The area in each tail is 0.025.

d) Type in the first probability of **0.025** and the same for the second probability. Click [OK].

$$P(-1.960 < Z < +1.96) = 0.9500.$$



Case 5 Dialog box



Graph window

6–2 Applications of the Normal Distribution

OBJECTIVE 4

Find probabilities for a normally distributed variable by transforming it into a standard normal variable.

The standard normal distribution curve can be used to solve a wide variety of practical problems. The only requirement is that the variable be normally or approximately normally distributed. There are several mathematical tests to determine whether a variable is normally distributed. See the Critical Thinking Challenges on page 366. For all the problems presented in this chapter, you can assume that the variable is normally or approximately normally distributed.

To solve problems by using the standard normal distribution, transform the original variable to a standard normal distribution variable by using the formula

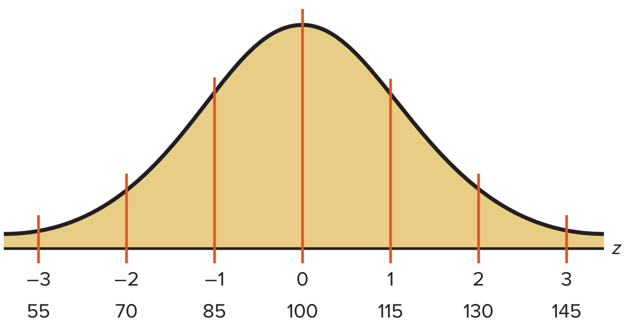
$$z = \frac{\text{value} - \text{mean}}{\text{standard deviation}} \qquad \text{or} \qquad z = \frac{X - \mu}{\sigma}$$

This is the same formula presented in Section 3–3. This formula transforms the values of the variable into standard units or z values. Once the variable is transformed, then the Procedure Table and Table E in Appendix A can be used to solve problems.

For example, suppose that the scores for a standardized test are normally distributed, have a mean of 100, and have a standard deviation of 15. When the scores are transformed to z values, the two distributions coincide, as shown in Figure 6–17. (Recall that the z distribution has a mean of 0 and a standard deviation of 1.)

FIGURE 6–17

Test Scores and Their Corresponding z Values



Note: The z values are rounded to two decimal places because Table E gives the z values to two decimal places.

Since we now have the ability to find the area under the standard normal curve, we can find the area under any normal curve by transforming the values of the variable to z values, and then we find the areas under the standard normal distribution, as shown in Section 6–1.

This procedure is summarized next.

Procedure Table

Finding the Area Under Any Normal Curve

Step 1 Draw a normal curve and shade the desired area.

Step 2 Convert the values of X to z values, using the formula $z = \dfrac{X - \mu}{\sigma}$.

Step 3 Find the corresponding area, using a table, calculator, or software.

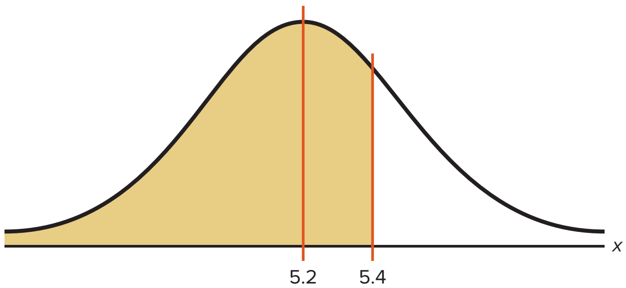
EXAMPLE 6–6 Liters of Blood

An adult has on average 5.2 liters of blood. Assume the variable is normally distributed and has a standard deviation of 0.3. Find the percentage of people who have less than 5.4 liters of blood in their system.

SOLUTION

Step 1 Draw a normal curve and shade the desired area. See Figure 6–18.

FIGURE 6–18
Area Under a Normal Curve for Example 6–6

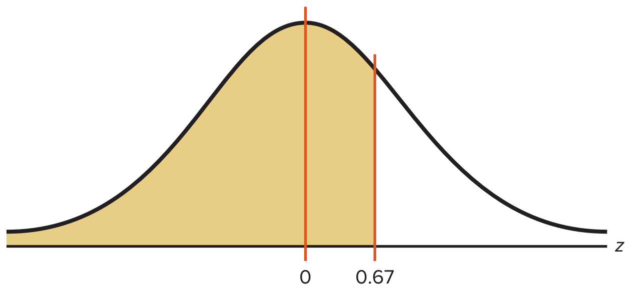


Step 2 Find the z value corresponding to 5.4.

$$z = \frac{X - \mu}{\sigma} = \frac{5.4 - 5.2}{0.3} = \frac{0.2}{0.3} = 0.67$$

Hence, 5.4 is 0.67 of a standard deviation above the mean, as shown in Figure 6–19.

FIGURE 6–19
Area and z Values for Example 6–6



Step 3 Find the corresponding area in Table E. The area under the standard normal curve to the left of $z = 0.67$ is 0.7486.

Therefore, 0.7486, or 74.86%, of adults have less than 5.4 liters of blood in their system.

EXAMPLE 6–7 Monthly Newspaper Recycling

Each month, an American household generates an average of 28 pounds of newspaper for garbage or recycling. Assume the variable is approximately normally distributed and the standard deviation is 2 pounds. If a household is selected at random, find the probability of its generating

a. Between 27 and 31 pounds per month

b. More than 30.2 pounds per month

Source: Michael D. Shook and Robert L. Shook, *The Book of Odds.*

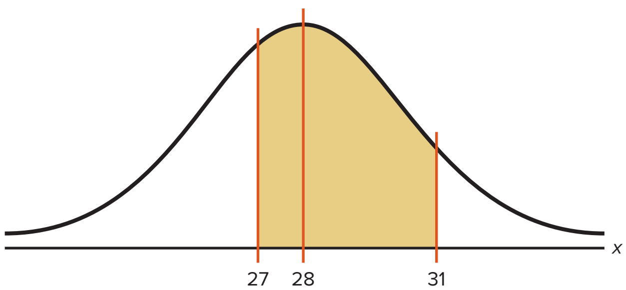
SOLUTION *a*

Step 1 Draw a normal curve and shade the desired area. See Figure 6–20.

Historical Note

Astronomers in the late 1700s and the 1800s used the principles underlying the normal distribution to correct measurement errors that occurred in charting the positions of the planets.

FIGURE 6–20
Area Under a Normal Curve for Part *a* of Example 6–7



Step 2 Find the two z values.

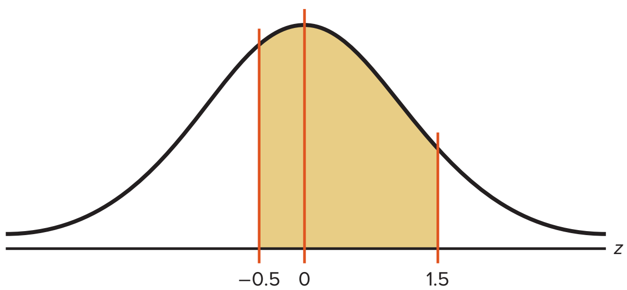
$$z_1 = \frac{X - \mu}{\sigma} = \frac{27 - 28}{2} = -\frac{1}{2} = -0.5$$

$$z_2 = \frac{X - \mu}{\sigma} = \frac{31 - 28}{2} = \frac{3}{2} = 1.5$$

Step 3 Find the appropriate area, using Table E. The area to the left of z_2 is 0.9332, and the area to the left of z_1 is 0.3085. Hence, the area between z_1 and z_2 is $0.9332 - 0.3085 = 0.6247$. See Figure 6–21.

© Nathan Mead/age fotostock RF

FIGURE 6–21
Area and z Values for Part *a* of Example 6–7

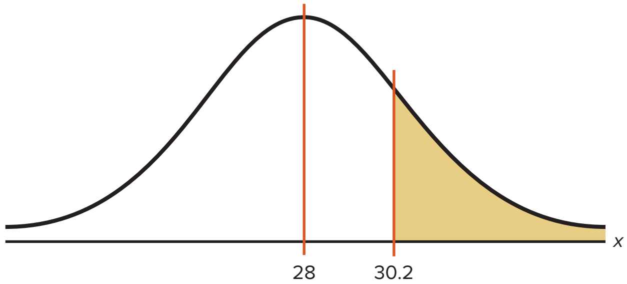


Hence, the probability that a randomly selected household generates between 27 and 31 pounds of newspapers per month is 62.47%.

SOLUTION *b*

Step 1 Draw a normal curve and shade the desired area, as shown in Figure 6–22.

FIGURE 6–22
Area Under a Normal Curve for Part *b* of Example 6–7

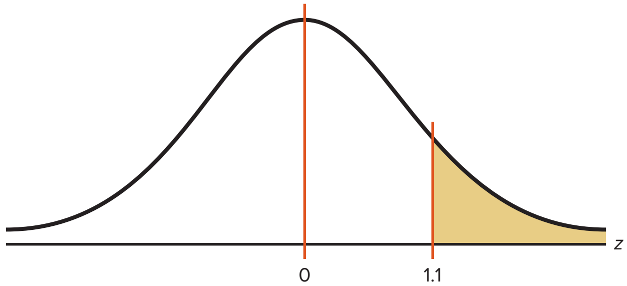


Step 2 Find the z value for 30.2.

$$z = \frac{X - \mu}{\sigma} = \frac{30.2 - 28}{2} = \frac{2.2}{2} = 1.1$$

Step 3 Find the appropriate area. The area to the left of $z = 1.1$ is 0.8643. Hence, the area to the right of $z = 1.1$ is $1.0000 - 0.8643 = 0.1357$. See Figure 6–23.

FIGURE 6–23
Area and z Values for Part *b* of Example 6–7



Hence, the probability that a randomly selected household will accumulate more than 30.2 pounds of newspapers is 0.1357, or 13.57%.

A normal distribution can also be used to answer questions of "How many?" This application is shown in Example 6–8.

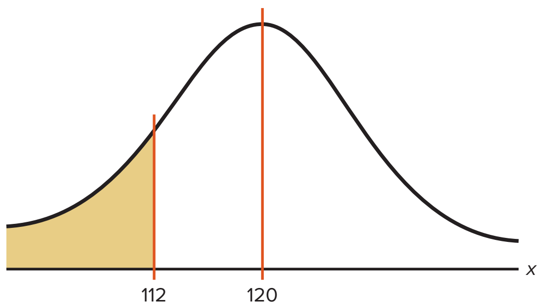
EXAMPLE 6–8 Decibels at a Concert

A decibel is a measure of the intensity of sound. The average number of decibels at a full concert is 120. Assume that the variable is approximately normally distributed and the standard deviation is 6. If 100 concerts are selected, approximately how many will have a decibel level less than 112?

SOLUTION

Step 1 Draw a normal curve and shade in the desired area. See Figure 6–24.

FIGURE 6–24
Area Under a Normal Curve for Example 6–8

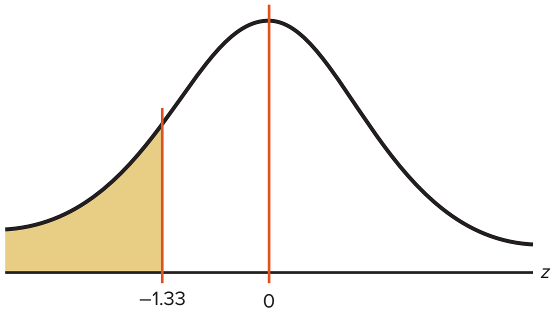


Step 2 Find the z value corresponding to 112.

$$z = \frac{X - \mu}{\sigma} = \frac{112 - 120}{6} = -1.33$$

Step 3 Find the area in Table E corresponding to $z = -1.33$. It is 0.0918. See Figure 6–25.

FIGURE 6–25
Area and z Values for Example 6–8



To find the number of concerts whose decibel output is less than 112, multiply $100 \times 0.0918 = 9.18$. Round this to 9. Hence, approximately 9 concerts out of 100 have a decibel output less than 112.

Note: For problems using percentages, be sure to write the percentage as a decimal before multiplying. Also round the answer to the nearest whole number, if necessary.

Finding Data Values Given Specific Probabilities

A normal distribution can also be used to find specific data values for given percentages. In this case, you are given a probability or percentage and need to find the corresponding data value X. You can use the formula $z = \frac{X - \mu}{\sigma}$. Substitute the values for z, μ, and σ; then solve for X. However, you can transform the formula and solve for X as shown.

$$z = \frac{X - \mu}{\sigma}$$

$$z \cdot \sigma = X - \mu \qquad \text{Multiply both sides by } \sigma.$$

$$z \cdot \sigma + \mu = X - \mu + \mu \qquad \text{Add } \mu \text{ to both sides.}$$

$$X = z \cdot \sigma + \mu \qquad \text{Exchange both sides of the equation.}$$

Now you can use this new formula and find the value for X.

Formula for Finding the Value of a Normal Variable *X*

$$X = z \cdot \sigma + \mu$$

The complete procedure for finding an X value is summarized in the Procedure Table shown.

Procedure Table

Finding Data Values for Specific Probabilities

Step 1 Draw a normal curve and shade the desired area that represents the probability, proportion, or percentile.

Step 2 Find the z value from the table that corresponds to the desired area.

Step 3 Calculate the X value by using the formula $X = z\sigma + \mu$.

OBJECTIVE 5

Find specific data values for given percentages, using the standard normal distribution.

EXAMPLE 6–9 Police Academy Qualifications

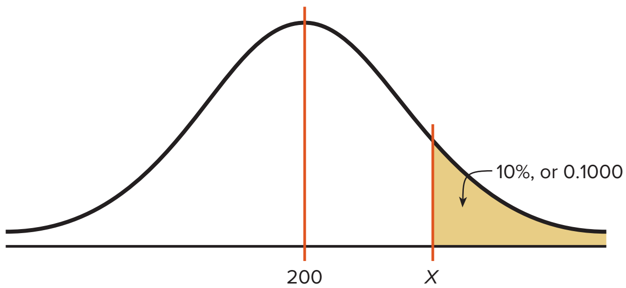
To qualify for a police academy, candidates must score in the top 10% on a general abilities test. Assume the test scores are normally distributed and the test has a mean of 200 and a standard deviation of 20. Find the lowest possible score to qualify.

SOLUTION

Step 1 Draw a normal distribution curve and shade the desired area that represents the probability.

Since the test scores are normally distributed, the test value X that cuts off the upper 10% of the area under a normal distribution curve is desired. This area is shown in Figure 6–26.

FIGURE 6–26
Area Under a Normal Curve for Example 6–9



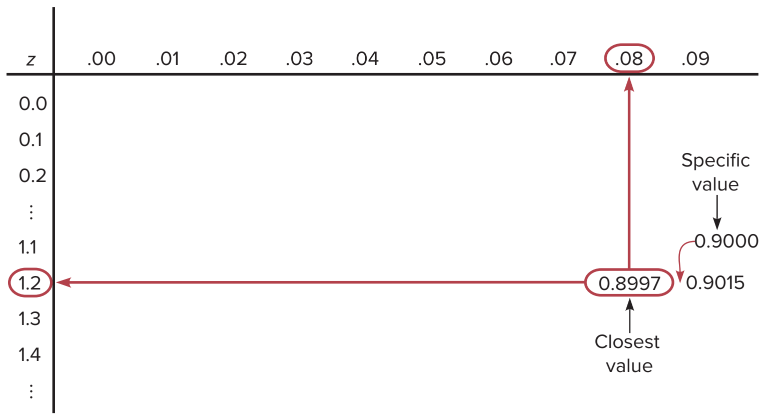
Work backward to solve this problem.

Subtract 0.1000 from 1.0000 to get the area under the normal distribution to the left of x: $1.0000 - 0.1000 = 0.9000$.

Step 2 Find the z value from Table E that corresponds to the desired area.

Find the z value that corresponds to an area of 0.9000 by looking up 0.9000 in the area portion of Table E. If the specific value cannot be found, use the closest value—in this case 0.8997, as shown in Figure 6–27. The corresponding z value is 1.28. (If the area falls exactly halfway between two z values, use the larger of the two z values. For example, the area 0.9500 falls halfway between 0.9495 and 0.9505. In this case use 1.65 rather than 1.64 for the z value.)

FIGURE 6–27
Finding the z Value from Table E (Example 6–9)



Step 3 Find the X value.

$$X = z \cdot \sigma + \mu = 1.28(20) + 200 = 25.6 + 200$$
$$= 225.6 = 226 \text{ (rounded)}$$

A score of 226 should be used as a cutoff. Anybody scoring 226 or higher qualifies for the academy.

Interesting Fact

Americans are the largest consumers of chocolate. We spend $16.6 billion annually.

EXAMPLE 6–10 Systolic Blood Pressure

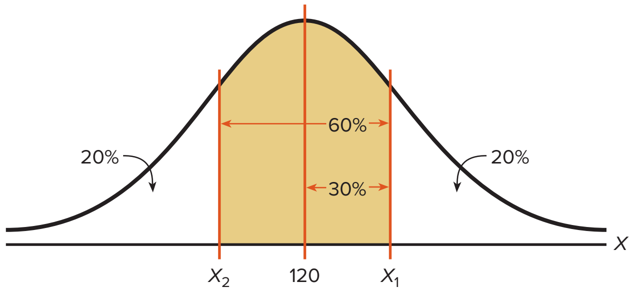
For a medical study, a researcher wishes to select people in the middle 60% of the population based on blood pressure. Assuming that blood pressure readings are normally distributed and the mean systolic blood pressure is 120 and the standard deviation is 8, find the upper and lower readings that would qualify people to participate in the study.

SOLUTION

Step 1 Draw a normal distribution curve and shade the desired area. The cutoff points are shown in Figure 6–28.

Two values are needed, one above the mean and one below the mean.

FIGURE 6–28
Area Under a Normal Curve for Example 6–10



Step 2 Find the z values.

To get the area to the left of the positive z value, add $0.5000 + 0.3000 = 0.8000$ ($30\% = 0.3000$). The z value with area to the left closest to 0.8000 is 0.84.

Step 3 Calculate the X values.

Substituting in the formula $X = z\sigma + \mu$ gives

$$X_1 = z\sigma + \mu = (0.84)(8) + 120 = 126.72$$

The area to the left of the negative z value is 20%, or 0.2000. The area closest to 0.2000 is −0.84.

$$X_2 = (-0.84)(8) + 120 = 113.28$$

Therefore, the middle 60% will have blood pressure readings of $113.28 < X < 126.72$.

As shown in this section, a normal distribution is a useful tool in answering many questions about variables that are normally or approximately normally distributed.

Determining Normality

A normally shaped or bell-shaped distribution is only one of many shapes that a distribution can assume; however, it is very important since many statistical methods require that the distribution of values (shown in subsequent chapters) be normally or approximately normally shaped.

There are several ways statisticians check for normality. The easiest way is to draw a histogram for the data and check its shape. If the histogram is not approximately bell-shaped, then the data are not normally distributed.

Skewness can be checked by using the Pearson coefficient (PC) of skewness also called Pearson's index of skewness. The formula is

$$PC = \frac{3(\overline{X} - \text{median})}{s}$$

If the index is greater than or equal to +1 or less than or equal to −1, it can be concluded that the data are significantly skewed.

In addition, the data should be checked for outliers by using the method shown in Chapter 3. Even one or two outliers can have a big effect on normality.

Examples 6–11 and 6–12 show how to check for normality.

EXAMPLE 6–11 Technology Inventories

A survey of 18 high-tech firms showed the number of days' inventory they had on hand. Determine if the data are approximately normally distributed.

5	29	34	44	45	63	68	74	74
81	88	91	97	98	113	118	151	158

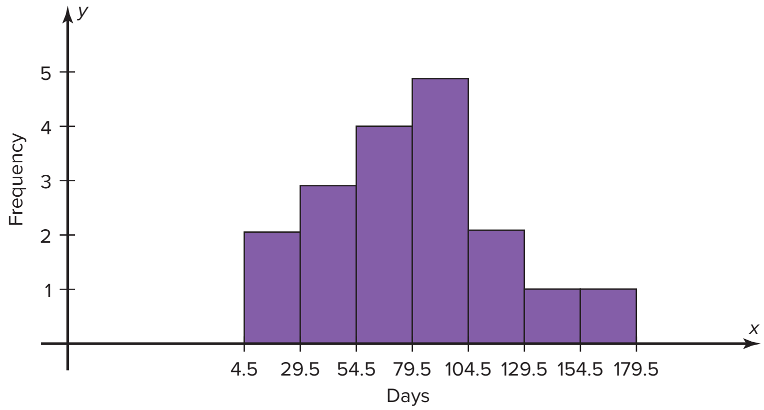
Source: USA TODAY.

SOLUTION

Step 1 Construct a frequency distribution and draw a histogram for the data, as shown in Figure 6–29.

Class	Frequency
5–29	2
30–54	3
55–79	4
80–104	5
105–129	2
130–154	1
155–179	1

FIGURE 6–29
Histogram for Example 6–11



Since the histogram is approximately bell-shaped, we can say that the distribution is approximately normal.

Step 2 Check for skewness. For these data, $\overline{X} = 79.5$, median = 77.5, and $s = 40.5$. Using the Pearson coefficient of skewness gives

$$\text{PC} = \frac{3(79.5 - 77.5)}{40.5}$$

$$= 0.148$$

In this case, PC is not greater than +1 or less than −1, so it can be concluded that the distribution is not significantly skewed.

Step 3 Check for outliers. Recall that an outlier is a data value that lies more than 1.5(IQR) units below Q_1 or 1.5(IQR) units above Q_3. In this case, $Q_1 = 45$ and $Q_3 = 98$; hence, IQR $= Q_3 - Q_1 = 98 - 45 = 53$. An outlier would be a data value less than $45 - 1.5(53) = -34.5$ or a data value larger than $98 + 1.5(53) = 177.5$. In this case, there are no outliers.

Since the histogram is approximately bell-shaped, the data are not significantly skewed, and there are no outliers, it can be concluded that the distribution is approximately normally distributed.

EXAMPLE 6–12 Number of Baseball Games Played

The data shown consist of the number of games played each year in the career of Baseball Hall of Famer Bill Mazeroski. Determine if the data are approximately normally distributed.

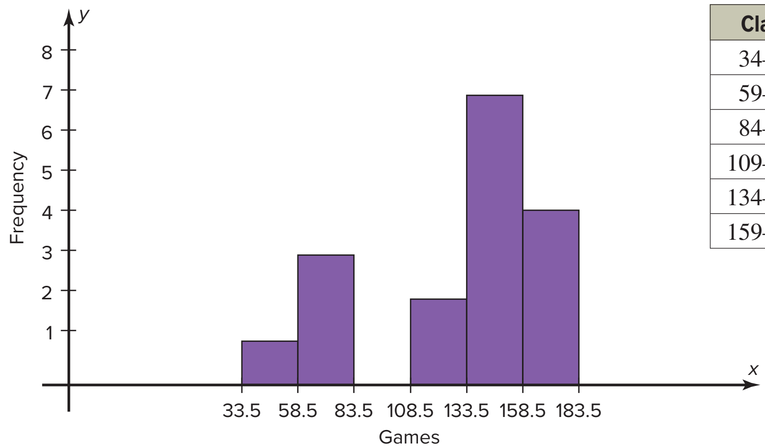
81	148	152	135	151	152
159	142	34	162	130	162
163	143	67	112	70	

Source: Greensburg Tribune Review.

SOLUTION

Step 1 Construct a frequency distribution and draw a histogram for the data. See Figure 6–30.

FIGURE 6–30 Histogram for Example 6–12



Class	Frequency
34–58	1
59–83	3
84–108	0
109–133	2
134–158	7
159–183	4

The histogram shows that the frequency distribution is somewhat negatively skewed.

Unusual Stats

The average amount of money stolen by a pickpocket each time is $128.

Step 2 Check for skewness; $\overline{X} = 127.24$, median $= 143$, and $s = 39.87$.

$$\text{PC} = \frac{3(\overline{X} - \text{median})}{s}$$
$$= \frac{3(127.24 - 143)}{39.87}$$
$$\approx -1.19$$

Since the PC is less than -1, it can be concluded that the distribution is significantly skewed to the left.

Step 3 Check for outliers. In this case, $Q_1 = 96.5$ and $Q_3 = 155.5$. IQR = $Q_3 - Q_1 = 155.5 - 96.5 = 59$. Any value less than $96.5 - 1.5(59) = 8$ or above $155.5 + 1.5(59) = 244$ is considered an outlier. There are no outliers.

In summary, the distribution is somewhat negatively skewed.

Another method that is used to check normality is to draw a *normal quantile plot. Quantiles,* sometimes called *fractiles,* are values that separate the data set into approximately equal groups. Recall that quartiles separate the data set into four approximately equal groups, and deciles separate the data set into 10 approximately equal groups. A normal quantile plot consists of a graph of points using the data values for the x coordinates and the z values of the quantiles corresponding to the x values for the y coordinates. (*Note:* The calculations of the z values are somewhat complicated, and technology is usually used to draw the graph. The Technology Step by Step section shows how to draw a normal quantile plot.) If the points of the quantile plot do not lie in an approximately straight line, then normality can be rejected.

There are several other methods used to check for normality. A method using normal probability graph paper is shown in the Critical Thinking Challenge section at the end of this chapter, and the chi-square goodness-of-fit test is shown in Chapter 11. Two other tests sometimes used to check normality are the Kolmogorov-Smirnov test and the Lilliefors test. An explanation of these tests can be found in advanced texts.

Applying the Concepts 6–2

Smart People

Assume you are thinking about starting a Mensa chapter in your hometown, which has a population of about 10,000 people. You need to know how many people would qualify for Mensa, which requires an IQ of at least 130. You realize that IQ is normally distributed with a mean of 100 and a standard deviation of 15. Complete the following.

1. Find the approximate number of people in your hometown who are eligible for Mensa.
2. Is it reasonable to continue your quest for a Mensa chapter in your hometown?
3. How could you proceed to find out how many of the eligible people would actually join the new chapter? Be specific about your methods of gathering data.
4. What would be the minimum IQ score needed if you wanted to start an Ultra-Mensa club that included only the top 1% of IQ scores?

See page 368 for the answers.

Exercises 6–2

1. Automobile Workers A worker in the automobile industry works an average of 43.7 hours per week. If the distribution is approximately normal with a standard deviation of 1.6 hours, what is the probability that a randomly selected automobile worker works less than 40 hours per week?

2. Teachers' Salaries The average annual salary for all U.S. teachers is \$47,750. Assume that the distribution is normal and the standard deviation is \$5680. Find the probability that a randomly selected teacher earns

a. Between \$35,000 and \$45,000 a year

b. More than \$40,000 a year

c. If you were applying for a teaching position and were offered \$31,000 a year, how would you feel (based on this information)?

Source: New York Times Almanac.

3. Population in U.S. Jails The average daily jail population in the United States is 706,242. If the distribution is normal and the standard deviation is 52,145, find the probability that on a randomly selected day, the jail population is

a. Greater than 750,000

b. Between 600,000 and 700,000

Source: New York Times Almanac.

4. **SAT Scores** The national average SAT score (for Verbal and Math) is 1028. If we assume a normal distribution with $\sigma = 92$, what is the 90th percentile score? What is the probability that a randomly selected score exceeds 1200?

Source: New York Times Almanac.

5. **Chocolate Bar Calories** The average number of calories in a 1.5-ounce chocolate bar is 225. Suppose that the distribution of calories is approximately normal with $\sigma = 10$. Find the probability that a randomly selected chocolate bar will have
   a. Between 200 and 220 calories
   b. Less than 200 calories

Source: The Doctor's Pocket Calorie, Fat, and Carbohydrate Counter.

6. **Monthly Mortgage Payments** The average monthly mortgage payment including principal and interest is \$982 in the United States. If the standard deviation is approximately \$180 and the mortgage payments are approximately normally distributed, find the probability that a randomly selected monthly payment is
   a. More than \$1000
   b. More than \$1475
   c. Between \$800 and \$1150

Source: World Almanac.

7. **Prison Sentences** The average prison sentence for a person convicted of second-degree murder is 15 years. If the sentences are normally distributed with a standard deviation of 2.1 years, find these probabilities:
   a. A prison sentence is greater than 18 years.
   b. A prison sentence is less than 13 years.

Source: The Book of Risks.

8. **Doctoral Student Salaries** Full-time Ph.D. students receive an average of \$12,837 per year. If the average salaries are normally distributed with a standard deviation of \$1500, find these probabilities.
   a. The student makes more than \$15,000.
   b. The student makes between \$13,000 and \$14,000.

Source: U.S. Education Dept., *Chronicle of Higher Education.*

9. **Miles Driven Annually** The mean number of miles driven per vehicle annually in the United States is 12,494 miles. Choose a randomly selected vehicle, and assume the annual mileage is normally distributed with a standard deviation of 1290 miles. What is the probability that the vehicle was driven more than 15,000 miles? Less than 8000 miles? Would you buy a vehicle if you had been told that it had been driven less than 6000 miles in the past year?

Source: World Almanac.

10. **Commute Time to Work** The average commute to work (one way) is 25 minutes according to the 2005 American Community Survey. If we assume that commuting times are normally distributed and that the standard deviation is 6.1 minutes, what is the probability that a randomly selected commuter spends more than 30 minutes commuting one way? Less than 18 minutes?

Source: www.census.gov

11. **Credit Card Debt** The average credit card debt for college seniors is \$3262. If the debt is normally distributed with a standard deviation of \$1100, find these probabilities.
   a. The senior owes at least \$1000.
   b. The senior owes more than \$4000.
   c. The senior owes between \$3000 and \$4000.

Source: USA TODAY.

12. **Price of Gasoline** The average retail price of gasoline (all types) for 2014 was 342 cents. What would the standard deviation have to be in order for there to be a 15% probability that a gallon of gas costs less than \$3.00?

Source: World Almanac.

13. **Potholes** The average number of potholes per 10 miles of paved U.S. roads is 130. Assume this variable is approximately normally distributed and has a standard deviation of 5. Find the probability that a randomly selected road has
   a. More than 142 potholes per 10 miles
   b. Less than 125 potholes per 10 miles
   c. Between 128 and 136 potholes per 10 miles

Source: Infrastructure Technology Institute.

14. **Newborn Elephant Weights** Newborn elephant calves usually weigh between 200 and 250 pounds—until October 2006, that is. An Asian elephant at the Houston (Texas) Zoo gave birth to a male calf weighing in at a whopping 384 pounds! Mack (like the truck) is believed to be the heaviest elephant calf ever born at a facility accredited by the Association of Zoos and Aquariums. If, indeed, the mean weight for newborn elephant calves is 225 pounds with a standard deviation of 45 pounds, what is the probability of a newborn weighing at least 384 pounds? Assume that the weights of newborn elephants are normally distributed.

Source: www.houstonzoo.org

15. **Heart Rates** For a certain group of individuals, the average heart rate is 72 beats per minute. Assume the variable is normally distributed and the standard deviation is 3 beats per minute. If a subject is selected at random, find the probability that the person has the following heart rate.
   a. Between 68 and 74 beats per minute
   b. Higher than 70 beats per minute
   c. Less than 75 beats per minute

16. **Salary of Full Professors** The average salary of a male full professor at a public four-year institution offering classes at the doctoral level is \$99,685. For a female

full professor at the same kind of institution, the salary is $90,330. If the standard deviation for the salaries of both genders is approximately $5200 and the salaries are normally distributed, find the 80th percentile salary for male professors and for female professors.

Source: World Almanac.

17. Cat Behavior A report stated that the average number of times a cat returns to its food bowl during the day is 36. Assuming the variable is normally distributed with a standard deviation of 5, what is the probability that a cat would return to its dish between 32 and 38 times a day?

18. Itemized Charitable Contributions The average charitable contribution itemized per income tax return in Pennsylvania is $792. Suppose that the distribution of contributions is normal with a standard deviation of $103. Find the limits for the middle 50% of contributions.

Source: IRS, *Statistics of Income Bulletin.*

19. New Home Sizes A contractor decided to build homes that will include the middle 80% of the market. If the average size of homes built is 1810 square feet, find the maximum and minimum sizes of the homes the contractor should build. Assume that the standard deviation is 92 square feet and the variable is normally distributed.

Source: Michael D. Shook and Robert L. Shook, *The Book of Odds.*

20. New-Home Prices If the average price of a new one-family home is $246,300 with a standard deviation of $15,000, find the minimum and maximum prices of the houses that a contractor will build to satisfy the middle 80% of the market. Assume that the variable is normally distributed.

Source: New York Times Almanac.

21. Cost of Personal Computers The average price of a personal computer (PC) is $949. If the computer prices are approximately normally distributed and $\sigma = \$100$, what is the probability that a randomly selected PC costs more than $1200? The least expensive 10% of personal computers cost less than what amount?

Source: New York Times Almanac.

22. Reading Improvement Program To help students improve their reading, a school district decides to implement a reading program. It is to be administered to the bottom 5% of the students in the district, based on the scores on a reading achievement exam. If the average score for the students in the district is 122.6, find the cutoff score that will make a student eligible for the program. The standard deviation is 18. Assume the variable is normally distributed.

23. Qualifying Test Scores To qualify for a medical study, an applicant must have a systolic blood pressure in the 50% of the middle range. If the systolic blood pressure is normally distributed with a mean of 120 and a standard deviation of 4, find the upper and lower limits of blood pressure a person must have to qualify for the study.

Source: Charleston Post and Courier.

24. Ages of Amtrak Passenger Cars The average age of Amtrak passenger train cars is 19.4 years. If the distribution of ages is normal and 20% of the cars are older than 22.8 years, find the standard deviation.

Source: New York Times Almanac.

25. Lengths of Hospital Stays The average length of a hospital stay for all diagnoses is 4.8 days. If we assume that the lengths of hospital stays are normally distributed with a variance of 2.1, then 10% of hospital stays are longer than how many days? Thirty percent of stays are less than how many days?

Source: www.cdc.gov

26. High School Competency Test A mandatory competency test for high school sophomores has a normal distribution with a mean of 400 and a standard deviation of 100.

a. The top 3% of students receive $500. What is the minimum score you would need to receive this award?

b. The bottom 1.5% of students must go to summer school. What is the minimum score you would need to stay out of this group?

27. Product Marketing An advertising company plans to market a product to low-income families. A study states that for a particular area, the average income per family is $24,596 and the standard deviation is $6256. If the company plans to target the bottom 18% of the families based on income, find the cutoff income. Assume the variable is normally distributed.

28. Bottled Drinking Water Americans drank an average of 34 gallons of bottled water per capita in 2014. If the standard deviation is 2.7 gallons and the variable is normally distributed, find the probability that a randomly selected American drank more than 25 gallons of bottled water. What is the probability that the selected person drank between 28 and 30 gallons?

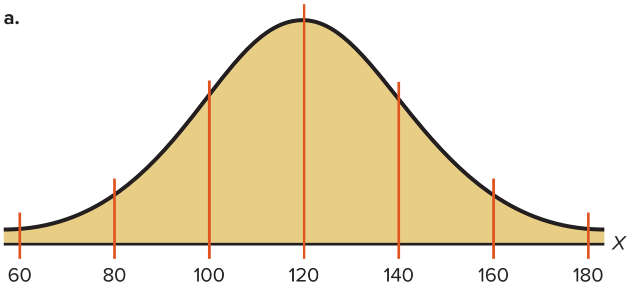
Source: Beverage Marketing Corporation

29. Wristwatch Lifetimes The mean lifetime of a wristwatch is 25 months, with a standard deviation of 5 months. If the distribution is normal, for how many months should a guarantee be made if the manufacturer does not want to exchange more than 10% of the watches? Assume the variable is normally distributed.

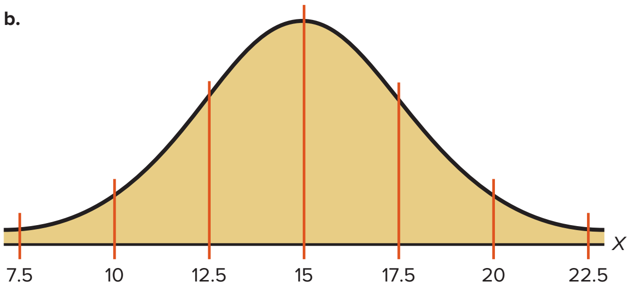
30. Police Academy Acceptance Exams To qualify for a police academy, applicants are given a test of physical fitness. The scores are normally distributed with a mean of 64 and a standard deviation of 9. If only the top 20% of the applicants are selected, find the cutoff score.

31. In the distributions shown, state the mean and standard deviation for each. *Hint:* See Figures 6–4 and 6–6. Also the vertical lines are 1 standard deviation apart.

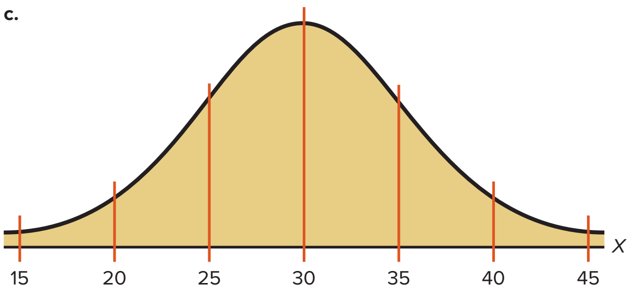
a.



b.



c.



32. SAT Scores Suppose that the mathematics SAT scores for high school seniors for a specific year have a mean of 456 and a standard deviation of 100 and are approximately normally distributed. If a subgroup of these high school seniors, those who are in the National Honor Society, is selected, would you expect the distribution of scores to have the same mean and standard deviation? Explain your answer.

33. Temperatures for Pittsburgh The mean temperature for a July in Pittsburgh is 73°. Assuming a normal distribution, what would the standard deviation have to be if 5% of the days have a temperature of at least 85°?

Source: The World Almanac.

34. Standardizing If a distribution of raw scores were plotted and then the scores were transformed to z scores, would the shape of the distribution change? Explain your answer.

35. Social Security Payments Consider the distribution of monthly Social Security (OASDI) payments. Assume a normal distribution with a standard deviation of $120. If one-fourth of the payments are above $1255.94, what is the mean monthly payment?

Source: World Almanac 2012.

36. Find the Mean In a normal distribution, find μ when σ is 6 and 3.75% of the area lies to the left of 85.

37. Internet Users U.S. internet users spend an average of 18.3 hours a week online. If 95% of users spend between 13.1 and 23.5 hours a week, what is the probability that a randomly selected user is online less than 15 hours a week?

Source: World Almanac 2012.

38. Exam Scores An instructor gives a 100-point examination in which the grades are normally distributed. The mean is 60 and the standard deviation is 10. If there are 5% A's and 5% F's, 15% B's and 15% D's, and 60% C's, find the scores that divide the distribution into those categories.

39. Drive-in Movies The data shown represent the number of outdoor drive-in movies in the United States for a 14-year period. Check for normality.

2084	1497	1014	910	899	870	837	859
848	826	815	750	637	737		

Source: National Association of Theater Owners.

40. Cigarette Taxes The data shown represent the cigarette tax (in cents) for the 50 U.S. states. Check for normality.

200	160	156	200	30	300	224	346	170	55
160	170	270	60	57	80	37	153	200	60
100	178	302	84	251	125	44	435	79	166
68	37	153	252	300	141	57	42	134	136
200	98	45	118	200	87	103	250	17	62

Source: http://www.tobaccofreekids.org

41. Box Office Revenues The data shown represent the box office total revenue (in millions of dollars) for a randomly selected sample of the top-grossing films in 2009. Check for normality.

37	32	155	277
146	80	66	113
71	29	166	36
28	72	32	32
30	32	52	84
37	402	42	109

Source: http://boxofficemojo.com

42. Number of Runs Made The data shown represent the number of runs made each year during Bill Mazeroski's career. Check for normality.

30	59	69	50	58	71	55	43	66	52	56	62
36	13	29	17	3							

Source: Greensburg Tribune Review.

43. Use your calculator to generate 20 random integers from 1–100, and check the set of data for normality. Would you expect these data to be normal? Explain.

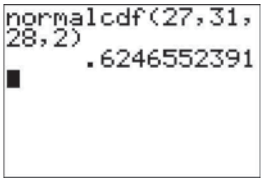
Technology Step by Step

TI-84 Plus Step by Step

Normal Random Variables

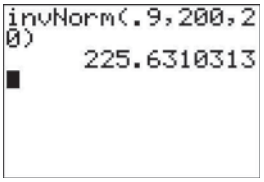
To find the probability for a normal random variable:
Press **2nd [DISTR],** then **2** for normalcdf(
The form is normalcdf(lower x value, upper x value, μ, σ)
Use E99 for ∞ (infinity) and $-$E99 for $-\infty$ (negative infinity). Press **2nd [EE]** to get E.

Example: Find the probability that x is between 27 and 31 when $\mu = 28$ and $\sigma = 2$.
(Example 6–7*a* from the text).
normalcdf(27,31,28,2)



To find the percentile for a normal random variable:
Press **2nd [DISTR],** then **3** for invNorm(
The form is invNorm(area to the left of x value, μ, σ)

Example: Find the 90th percentile when $\mu = 200$ and $\sigma = 20$ (Example 6–9 from text).
invNorm(.9,200,20)

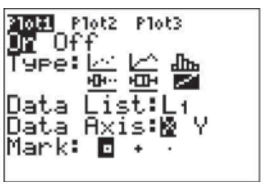


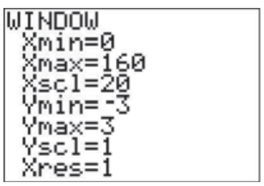
To construct a normal quantile plot:

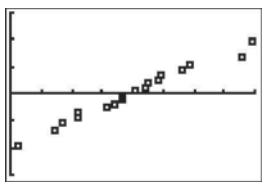
1. Enter the data values into L_1.
2. Press **2nd [STAT PLOT]** to get the STAT PLOT menu.
3. Press **1** for Plot 1.
4. Turn on the plot by pressing **ENTER** while the cursor is flashing over ON.
5. Move the cursor to the normal quantile plot (6th graph).
6. Make sure L_1 is entered for the Data List and X is highlighted for the Data Axis.
7. Press **WINDOW** for the Window menu. Adjust Xmin and Xmax according to the data values. Adjust Ymin and Ymax as well; Ymin = -3 and Ymax = 3 usually work fine.
8. Press **GRAPH.**

Example: A data set is given below. Check for normality by constructing a normal quantile plot.

5	29	34	44	45	63	68	74	74
81	88	91	97	98	113	118	151	158







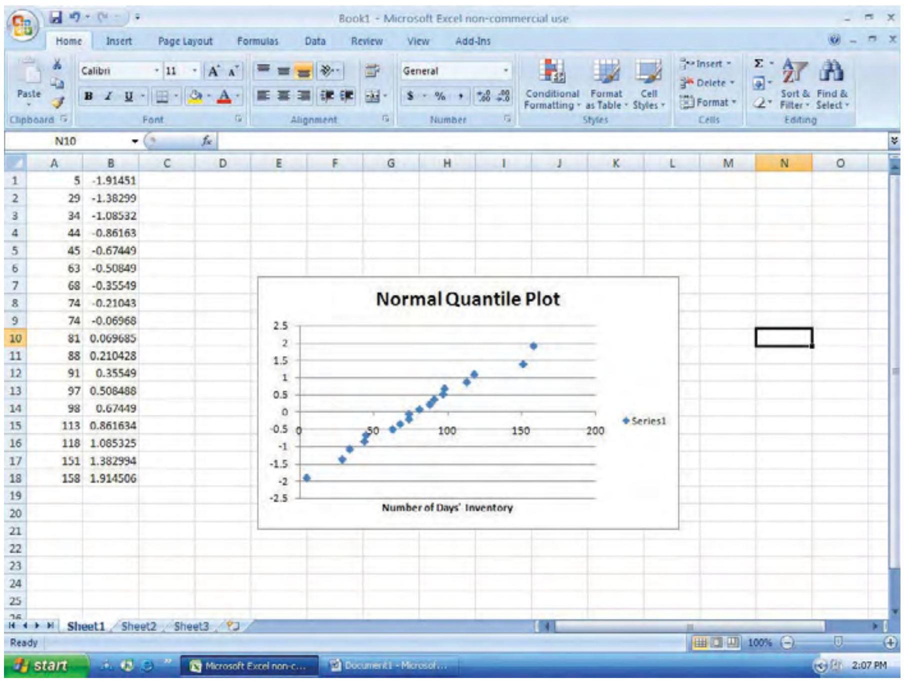
Since the points in the normal quantile plot lie close to a straight line, the distribution is approximately normal.

EXCEL
Step by Step

Normal Quantile Plot

Excel can be used to construct a normal quantile plot in order to examine if a set of data is approximately normally distributed.

1. Enter the data from the MINITAB Example 6–1 (see next page) into column A of a new worksheet. The data should be sorted in ascending order. If the data are not already sorted in ascending order, highlight the data to be sorted and select the Sort & Filter icon from the toolbar. Then select Sort Smallest to Largest.
2. After all the data are entered and sorted in column A, select cell B1. Type: =NORMSINV(1/(2*18)). Since the sample size is 18, each score represents $\frac{1}{18}$, or approximately 5.6%, of the sample. Each data value is assumed to subdivide the data into equal intervals. Each data value corresponds to the midpoint of a particular subinterval. Thus, this procedure will standardize the data by assuming each data value represents the midpoint of a subinterval of width $\frac{1}{18}$.
3. Repeat the procedure from step 2 for each data value in column A. However, for each subsequent value in column A, enter the next odd multiple of $\frac{1}{36}$ in the argument for the NORMSINV function. For example, in cell B2, type: =NORMSINV(3/(2*18)). In cell B3, type: =NORMSINV(5/(2*18)), and so on until all the data values have corresponding z scores.
4. Highlight the data from columns A and B, and select Insert, then Scatter chart. Select the Scatter with only markers (the first Scatter chart).
5. To insert a title to the chart: Left-click on any region of the chart. Select Chart Tools and Layout from the toolbar. Then select Chart Title.
6. To insert a label for the variable on the horizontal axis: Left-click on any region of the chart. Select Chart Tools and Layout from the toolbar. Then select Axis Titles>Primary Horizontal Axis Title.



The points on the chart appear to lie close to a straight line. Thus, we deduce that the data are approximately normally distributed.

MINITAB
Step by Step

Determining Normality

There are several ways in which statisticians test a data set for normality. Four are shown here.

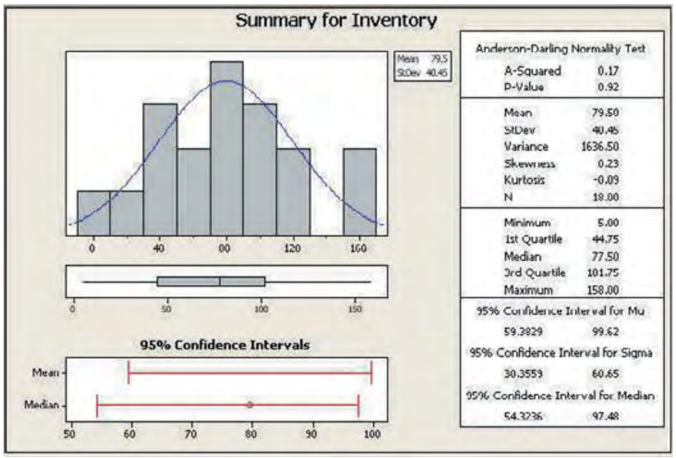
Data for Example 6–1

5	29	34	44	45
63	68	74	74	81
88	91	97	98	113
118	151	158		

Construct a Histogram

Inspect the histogram for shape.

1. Enter the data in the first column of a new worksheet. Name the column Inventory.
2. Use Stat>Basic Statistics>Graphical Summary to create the histogram. Is it symmetric? Is there a single peak? The instructions in Section 2–2 can be used to change the *X* scale to match the histogram.



Check for Outliers

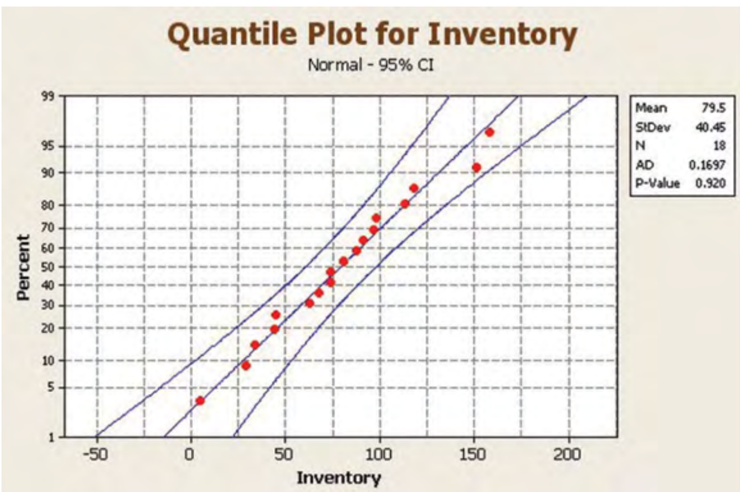
Inspect the boxplot for outliers. There are no outliers in this graph. Furthermore, the box is in the middle of the range, and the median is in the middle of the box. Most likely this is not a skewed distribution either.

Calculate the Pearson Coefficient of Skewness

The measure of skewness in the graphical summary is not the same as the Pearson coefficient. Use the calculator and the formula.

$$PC = \frac{3(\overline{X} - \text{median})}{s}$$

3. Select **Calc>Calculator,** then type **PC** in the text box for Store result in:.
4. Enter the expression: **3*(MEAN(C1)-MEDIAN(C1))/(STDEV(C1)).** Make sure you get all the parentheses in the right place!
5. Click [OK]. The result, 0.148318, will be stored in the first row of C2 named PC. Since it is smaller than +1, the distribution is not skewed.



Construct a Normal Probability Plot

6. Select **Graph>Probability Plot,** then Single and click [OK].
7. Double-click C1 Inventory to select the data to be graphed.
8. Click [Distribution] and make sure that Normal is selected. Click [OK].
9. Click [Labels] and enter the title for the graph: **Quantile Plot for Inventory.** You may also put **Your Name** in the subtitle.

10. Click [OK] twice. Inspect the graph to see if the graph of the points is linear.

These data are nearly normal.

What do you look for in the plot?

a) An "S curve" indicates a distribution that is too thick in the tails, a uniform distribution, for example.

b) Concave plots indicate a skewed distribution.

c) If one end has a point that is extremely high or low, there may be outliers.

This data set appears to be nearly normal by every one of the four criteria!

6–3 The Central Limit Theorem

OBJECTIVE 6

Use the central limit theorem to solve problems involving sample means for large samples.

In addition to knowing how individual data values vary about the mean for a population, statisticians are interested in knowing how the means of samples of the same size taken from the same population vary about the population mean.

Distribution of Sample Means

Suppose a researcher selects a sample of 30 adult males and finds the mean of the measure of the triglyceride levels for the sample subjects to be 187 milligrams/deciliter. Then suppose a second sample is selected, and the mean of that sample is found to be 192 milligrams/deciliter. Continue the process for 100 samples. What happens then is that the mean becomes a random variable, and the sample means 187, 192, 184, . . . , 196 constitute a *sampling distribution of sample means.*

A **sampling distribution of sample means** is a distribution using the means computed from all possible random samples of a specific size taken from a population.

If the samples are randomly selected with replacement, the sample means, for the most part, will be somewhat different from the population mean μ. These differences are caused by sampling error.

Sampling error is the difference between the sample measure and the corresponding population measure due to the fact that the sample is not a perfect representation of the population.

When all possible samples of a specific size are selected with replacement from a population, the distribution of the sample means for a variable has two important properties, which are explained next.

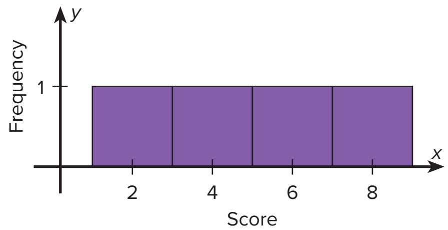
Properties of the Distribution of Sample Means

1. The mean of the sample means will be the same as the population mean.
2. The standard deviation of the sample means will be smaller than the standard deviation of the population, and it will be equal to the population standard deviation divided by the square root of the sample size.

The following example illustrates these two properties. Suppose a professor gave an 8-point quiz to a small class of four students. The results of the quiz were 2, 6, 4, and 8. For the sake of discussion, assume that the four students constitute the population. The mean of the population is

$$\mu = \frac{2 + 6 + 4 + 8}{4} = 5$$

FIGURE 6–31
Distribution of Quiz Scores



Historical Notes

Two mathematicians who contributed to the development of the central limit theorem were Abraham DeMoivre (1667–1754) and Pierre Simon Laplace (1749–1827). DeMoivre was once jailed for his religious beliefs. After his release, DeMoivre made a living by consulting on the mathematics of gambling and insurance. He wrote two books, *Annuities Upon Lives* and *The Doctrine of Chance*.

Laplace held a government position under Napoleon and later under Louis XVIII. He once computed the probability of the sun rising to be 18,226,214/18,226,215.

The standard deviation of the population is

$$\sigma = \frac{\sqrt{(2-5)^2 + (6-5)^2 + (4-5)^2 + (8-5)^2}}{4} \approx 2.236$$

The graph of the original distribution is shown in Figure 6–31. This is called a *uniform distribution.*

Now, if all samples of size 2 are taken with replacement and the mean of each sample is found, the distribution is as shown.

Sample	Mean	Sample	Mean
2, 2	2	6, 2	4
2, 4	3	6, 4	5
2, 6	4	6, 6	6
2, 8	5	6, 8	7
4, 2	3	8, 2	5
4, 4	4	8, 4	6
4, 6	5	8, 6	7
4, 8	6	8, 8	8

A frequency distribution of sample means is as follows.

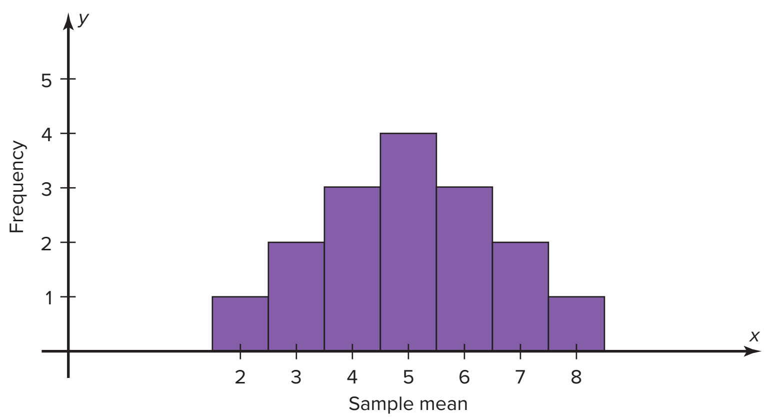
$\overline{X}$	f
2	1
3	2
4	3
5	4
6	3
7	2
8	1

For the data from the example just discussed, Figure 6–32 shows the graph of the sample means. The histogram appears to be approximately normal.

The mean of the sample means, denoted by $\mu_{\overline{X}}$, is

$$\mu_{\overline{X}} = \frac{2 + 3 + \cdots + 8}{16} = \frac{80}{16} = 5$$

FIGURE 6–32
Distribution of Sample Means



which is the same as the population mean. Hence,

$$\mu_{\overline{X}} = \mu$$

The standard deviation of sample means, denoted by $\sigma_{\overline{X}}$, is

$$\sigma_{\overline{X}} = \frac{\sqrt{(2-5)^2 + (3-5)^2 + \cdots + (8-5)^2}}{16} \approx 1.581$$

which is the same as the population standard deviation, divided by $\sqrt{2}$:

$$\sigma_{\overline{X}} = \frac{2.236}{\sqrt{2}} \approx 1.581$$

Unusual Stats

Each year a person living in the United States consumes on average 1400 pounds of food.

(*Note:* Rounding rules were not used here in order to show that the answers coincide.)

In summary, if all possible samples of size n are taken with replacement from the same population, the mean of the sample means, denoted by $\mu_{\overline{X}}$, equals the population mean μ; and the standard deviation of the sample means, denoted by $\sigma_{\overline{X}}$, equals $\sigma/\sqrt{n}$. The standard deviation of the sample means is called the **standard error of the mean.** Hence,

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

A third property of the sampling distribution of sample means pertains to the shape of the distribution and is explained by the **central limit theorem.**

The Central Limit Theorem

As the sample size n increases without limit, the shape of the distribution of the sample means taken with replacement from a population with mean μ and standard deviation σ will approach a normal distribution. As previously shown, this distribution will have a mean μ and a standard deviation $\sigma/\sqrt{n}$.

If the sample size is sufficiently large, the central limit theorem can be used to answer questions about sample means in the same manner that a normal distribution can be used to answer questions about individual values. The only difference is that a new formula must be used for the z values. It is

$$z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

Notice that $\overline{X}$ is the sample mean, and the denominator must be adjusted since means are being used instead of individual data values. The denominator is the standard deviation of the sample means.

If a large number of samples of a given size are selected from a normally distributed population, or if a large number of samples of a given size that is greater than or equal to 30 are selected from a population that is not normally distributed, and the sample means are computed, then the distribution of sample means will look like the one shown in Figure 6–33. Their percentages indicate the areas of the regions.

It's important to remember two things when you use the central limit theorem:

1. When the original variable is normally distributed, the distribution of the sample means will be normally distributed, for any sample size n.
2. When the distribution of the original variable is not normal, a sample size of 30 or more is needed to use a normal distribution to approximate the distribution of the sample means. The larger the sample, the better the approximation will be.

Examples 6–13 through 6–15 show how the standard normal distribution can be used to answer questions about sample means.

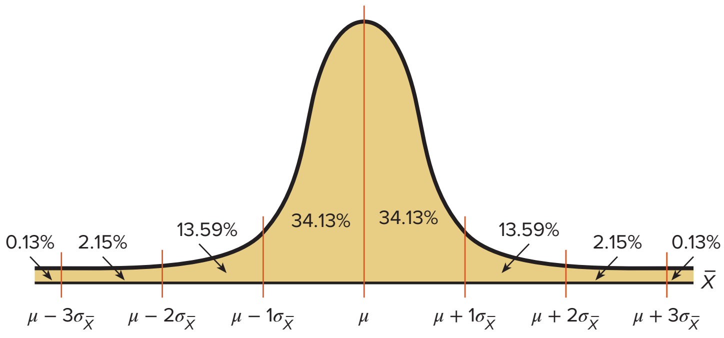


FIGURE 6–33
Distribution of Sample Means for a Large Number of Samples

EXAMPLE 6–13 Hours That Children Watch Television

A. C. Neilsen reported that children between the ages of 2 and 5 watch an average of 25 hours of television per week. Assume the variable is normally distributed and the standard deviation is 3 hours. If 20 children between the ages of 2 and 5 are randomly selected, find the probability that the mean of the number of hours they watch television will be greater than 26.3 hours.

Source: Michael D. Shook and Robert L. Shook, *The Book of Odds.*

SOLUTION

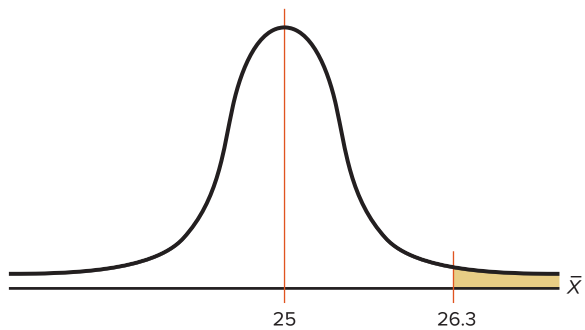
Since the variable is approximately normally distributed, the distribution of sample means will be approximately normal, with a mean of 25. The standard deviation of the sample means is

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{20}} = 0.671$$

Step 1 Draw a normal curve and shade the desired area.

The distribution of the means is shown in Figure 6–34, with the appropriate area shaded.

FIGURE 6–34
Distribution of the Means for Example 6–13



Step 2 Convert the $\overline{X}$ value to a z value.

The z value is

$$z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{26.3 - 25}{3/\sqrt{20}} = \frac{1.3}{0.671} = 1.94$$

Step 3 Find the corresponding area for the z value.

The area to the right of 1.94 is 1.000 − 0.9738 = 0.0262, or 2.62%.

One can conclude that the probability of obtaining a sample mean larger than 26.3 hours is 2.62% [that is, $P(\overline{X} > 26.3) = 0.0262$]. Specifically, the probability that the 20 children selected between the ages of 2 and 5 watch more than 26.3 hours of television per week is 2.62%.

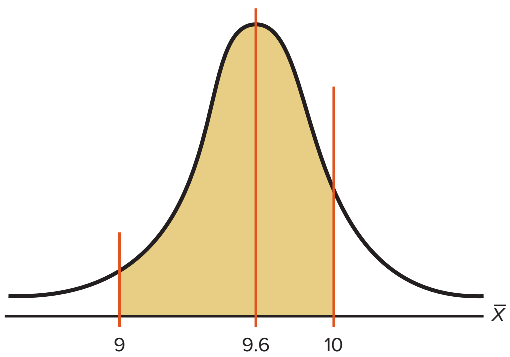
EXAMPLE 6–14 Drive Times

The average drive to work is 9.6 miles. Assume the standard deviation is 1.8 miles. If a random sample of 36 employed people who drive to work are selected, find the probability that the mean of the sample miles driven to work is between 9 and 10 miles.

SOLUTION

Step 1 Draw a normal curve and shade the desired area. Since the sample is 30 or larger, the normality assumption is not necessary. The desired area is shown in Figure 6–35.

FIGURE 6–35
Area Under a Normal Curve for Example 6–14



Step 2 Convert the $\overline{X}$ values to z values. The two z values are

$$z_1 = \frac{9 - 9.6}{1.8/\sqrt{36}} = -2$$

$$z_2 = \frac{10 - 9.6}{1.8/\sqrt{36}} = 1.33$$

Step 3 Find the corresponding area for each z value. To find the area between the two z values of -2 and 1.33, look up the corresponding areas in Table E and subtract the smaller area value from the larger area value. The area for $z = -2$ is 0.0228, and the area for $z = 1.33$ is 0.9082. Hence, the area between the two z values is $0.9082 - 0.0228 = 0.8854$, or 88.54%.

Hence, the probability of obtaining a sample mean between 9 and 10 miles is 88.54%; that is, $P(9 < \overline{X} < 10)$ is 0.8854. Specifically, the probability that the mean mileage driven to work for a sample size of 36 is between 9 and 10 miles is 88.54%.

Students sometimes have difficulty deciding whether to use

$$z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \qquad \text{or} \qquad z = \frac{X - \mu}{\sigma}$$

The formula

$$z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

should be used to gain information about a sample mean, as shown in this section. The formula

$$z = \frac{X - \mu}{\sigma}$$

is used to gain information about an individual data value obtained from the population. Notice that the first formula contains $\overline{X}$, the symbol for the sample mean, while the second formula contains X, the symbol for an individual data value. Example 6–15 illustrates the uses of the two formulas.

EXAMPLE 6–15 Working Weekends

The average time spent by construction workers who work on weekends is 7.93 hours (over 2 days). Assume the distribution is approximately normal and has a standard deviation of 0.8 hour.

a. Find the probability that an individual who works at that trade works fewer than 8 hours on the weekend.

b. If a sample of 40 construction workers is randomly selected, find the probability that the mean of the sample will be less than 8 hours.

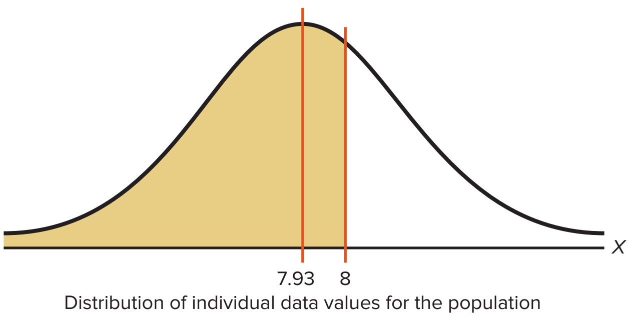
Source: Bureau of Labor Statistics.

SOLUTION *a*

Step 1 Draw a normal distribution and shade the desired area.

Since the question concerns an individual person, the formula $z = (X - \mu)/\sigma$ is used. The distribution is shown in Figure 6–36.

FIGURE 6–36
Area Under a Normal Curve for Part *a* of Example 6–15



Step 2 Find the z value.

$$Z = \frac{X - \mu}{\sigma} = \frac{8 - 7.93}{0.8} \approx 0.09$$

Step 3 Find the area to the left of $z = 0.09$.

It is 0.5359.

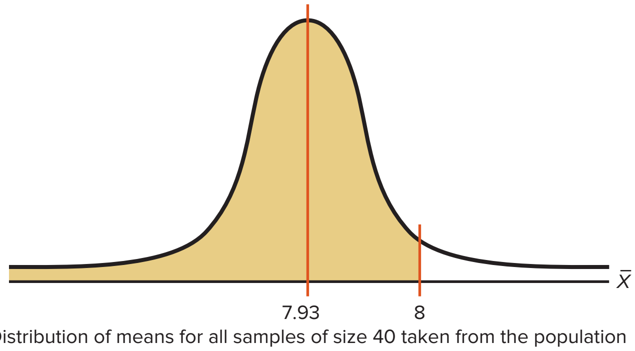
Hence, the probability of selecting a construction worker who works less than 8 hours on a weekend is 0.5359, or 53.59%.

SOLUTION b

Step 1 Draw a normal curve and shade the desired area.

Since the question concerns the mean of a sample with a size of 40, the central limit theorem formula $z = (\overline{X} - \mu)/(\sigma/\sqrt{n})$ is used. The area is shown in Figure 6–37.

FIGURE 6–37
Area Under a Normal Curve for Part *b* of Example 6–15



Step 2 Find the z value for a mean of 8 hours and a sample size of 40.

$$z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{8 - 7.93}{0.8/\sqrt{40}} \approx 0.55$$

Step 3 Find the area corresponding to $z = 0.55$. The area is 0.7088.

Hence, the probability of getting a sample mean of less than 8 hours when the sample size is 40 is 0.7088, or 70.88%.

Comparing the two probabilities, you can see the probability of selecting an individual construction worker who works less than 8 hours on a weekend is 53.59%. The probability of selecting a random sample of 40 construction workers with a mean of less than 8 hours per week is 70.88%. This difference of 17.29% is due to the fact that the distribution of sample means is much less variable than the distribution of individual data values. The reason is that as the sample size increases, the standard deviation of the means decreases.

Finite Population Correction Factor (Optional)

The formula for the standard error of the mean $\sigma/\sqrt{n}$ is accurate when the samples are drawn with replacement or are drawn without replacement from a very large or infinite population. Since sampling with replacement is for the most part unrealistic, a *correction factor* is necessary for computing the standard error of the mean for samples drawn without replacement from a finite population. Compute the correction factor by using the expression

$$\sqrt{\frac{N - n}{N - 1}}$$

where N is the population size and n is the sample size.

This correction factor is necessary if relatively large samples (usually greater than 5% of the population) are taken from a small population, because the sample mean will then more accurately estimate the population mean and there will be less error in the estimation. Therefore, the standard error of the mean must be multiplied by the correction factor to adjust for large samples taken from a small population. That is,

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N - n}{N - 1}}$$

Interesting Fact

The bubonic plague killed more than 25 million people in Europe between 1347 and 1351.

Finally, the formula for the z value becomes

$$z = \frac{\overline{X} - \mu}{\dfrac{\sigma}{\sqrt{n}} \cdot \sqrt{\dfrac{N-n}{N-1}}}$$

When the population is large and the sample is small, the correction factor is generally not used, since it will be very close to 1.00.

The formulas and their uses are summarized in Table 6–1.

TABLE 6–1 Summary of Formulas and Their Uses

Formula	Use
1. $z = \dfrac{X - \mu}{\sigma}$	Used to gain information about an individual data value when the variable is normally distributed
2. $z = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}$	Used to gain information when applying the central limit theorem about a sample mean when the variable is normally distributed or when the sample size is 30 or more

Applying the Concepts 6–3

Times To Travel to School

Twenty students from a statistics class each collected a random sample of times on how long it took students to get to class from their homes. All the sample sizes were 30. The resulting means are listed.

Student	Mean	Std. Dev.	Student	Mean	Std. Dev.
1	22	3.7	11	27	1.4
2	31	4.6	12	24	2.2
3	18	2.4	13	14	3.1
4	27	1.9	14	29	2.4
5	20	3.0	15	37	2.8
6	17	2.8	16	23	2.7
7	26	1.9	17	26	1.8
8	34	4.2	18	21	2.0
9	23	2.6	19	30	2.2
10	29	2.1	20	29	2.8

1. The students noticed that everyone had different answers. If you randomly sample over and over from any population, with the same sample size, will the results ever be the same?
2. The students wondered whose results were right. How can they find out what the population mean and standard deviation are?
3. Input the means into the computer and check if the distribution is normal.
4. Check the mean and standard deviation of the means. How do these values compare to the students' individual scores?
5. Is the distribution of the means a sampling distribution?
6. Check the sampling error for students 3, 7, and 14.
7. Compare the standard deviation of the sample of the 20 means. Is that equal to the standard deviation from student 3 divided by the square root of the sample size? How about for student 7, or 14?

See page 368 for the answers.

Exercises 6–3

1. If samples of a specific size are selected from a population and the means are computed, what is this distribution of means called?

2. Why do most of the sample means differ somewhat from the population mean? What is this difference called?

3. What is the mean of the sample means?

4. What is the standard deviation of the sample means called? What is the formula for this standard deviation?

5. What does the central limit theorem say about the shape of the distribution of sample means?

6. What formula is used to gain information about an individual data value when the variable is normally distributed?

For Exercises 7 through 25, assume that the sample is taken from a large population and the correction factor can be ignored.

7. Life of Smoke Detectors The average lifetime of smoke detectors that a company manufactures is 5 years, or 60 months, and the standard deviation is 8 months. Find the probability that a random sample of 30 smoke detectors will have a mean lifetime between 58 and 63 months.

8. Glass Garbage Generation A survey found that the American family generates an average of 17.2 pounds of glass garbage each year. Assume the standard deviation of the distribution is 2.5 pounds. Find the probability that the mean of a sample of 55 families will be between 17 and 18 pounds.

Source: Michael D. Shook and Robert L. Shook, *The Book of Odds.*

9. New Residences The average number of moves a person makes in his or her lifetime is 12. If the standard deviation is 3.2, find the probability that the mean of a sample of 36 people is

a. Less than 13

b. Greater than 13

c. Between 11 and 12

10. Teachers' Salaries in Connecticut The average teacher's salary in Connecticut (ranked first among states) is \$57,337. Suppose that the distribution of salaries is normal with a standard deviation of \$7500.

a. What is the probability that a randomly selected teacher makes less than \$52,000 per year?

b. If we sample 100 teachers' salaries, what is the probability that the sample mean is less than \$56,000?

Source: New York Times Almanac.

11. Earthquakes The average number of earthquakes that occur in Los Angeles over one month is 36. (Most are undetectable.) Assume the standard deviation is 3.6. If a random sample of 35 months is selected, find the probability that the mean of the sample is between 34 and 37.5.

Source: Southern California Earthquake Center.

12. Teachers' Salaries in North Dakota The average teacher's salary in North Dakota is \$37,764. Assume a normal distribution with $\sigma = \$5100$.

a. What is the probability that a randomly selected teacher's salary is greater than \$45,000?

b. For a sample of 75 teachers, what is the probability that the sample mean is greater than \$38,000?

Source: New York Times Almanac.

13. Movie Ticket Prices In the second quarter of 2015, the average movie ticket cost \$8.61. In a random sample of 50 movie tickets from various areas, what is the probability that the mean cost exceeds \$8.00, given that the population standard deviation is \$1.39?

Source: Variety.

14. SAT Scores The national average SAT score (for Verbal and Math) is 1028. Suppose that nothing is known about the shape of the distribution and that the standard deviation is 100. If a random sample of 200 scores were selected and the sample mean were calculated to be 1050, would you be surprised? Explain.

Source: New York Times Almanac.

15. Cost of Overseas Trip The average overseas trip cost is \$2708 per visitor. If we assume a normal distribution with a standard deviation of \$405, what is the probability that the cost for a randomly selected trip is more than \$3000? If we select a random sample of 30 overseas trips and find the mean of the sample, what is the probability that the mean is greater than \$3000?

Source: World Almanac.

16. Cell Phone Lifetimes A recent study of the lifetimes of cell phones found the average is 24.3 months. The standard deviation is 2.6 months. If a company provides its 33 employees with a cell phone, find the probability that the mean lifetime of these phones will be less than 23.8 months. Assume cell phone life is a normally distributed variable.

17. Water Use The *Old Farmer's Almanac* reports that the average person uses 123 gallons of water daily. If the standard deviation is 21 gallons, find the probability that the mean of a randomly selected sample of 15 people will be between 120 and 126 gallons. Assume the variable is normally distributed.

18. Medicare Hospital Insurance The average yearly Medicare Hospital Insurance benefit per person was \$4064 in a recent year. If the benefits are normally distributed with a standard deviation of \$460, find the probability that the mean benefit for a random sample of 20 patients is

a. Less than \$3800

b. More than \$4100

Source: New York Times Almanac.

19. Amount of Laundry Washed Each Year Procter & Gamble reported that an American family of four washes an average of 1 ton (2000 pounds) of clothes each year. If the standard deviation of the distribution is 187.5 pounds, find the probability that the mean of a randomly selected sample of 50 families of four will be between 1980 and 1990 pounds.

Source: The Harper's Index Book.

20. Per Capita Income of Delaware Residents In a recent year, Delaware had the highest per capita annual income with \$51,803. If $\sigma = \$4850$, what is the probability that a random sample of 34 state residents had a mean income greater than \$50,000? Less than \$48,000?

Source: New York Times Almanac.

21. Monthly Precipitation for Miami The mean precipitation for Miami in August is 8.9 inches. Assume that the standard deviation is 1.6 inches and the variable is normally distributed.

a. Find the probability that a randomly selected August month will have precipitation of less than 8.2 inches. (This month is selected from August months over the last 10 years.)

b. Find the probability that a sample of 10 August months will have a mean of less than 8.2 inches.

c. Does it seem reasonable that a randomly selected August month will have less than 8.2 inches of rain?

d. Does it seem reasonable that a sample of 10 months will have a mean of less than 8.2 months?

Source: National Climatic Data Center.

22. Systolic Blood Pressure Assume that the mean systolic blood pressure of normal adults is 120 millimeters of mercury (mm Hg) and the standard deviation is 5.6. Assume the variable is normally distributed.

a. If an individual is selected, find the probability that the individual's pressure will be between 120 and 121.8 mm Hg.

b. If a sample of 30 adults is randomly selected, find the probability that the sample mean will be between 120 and 121.8 mm Hg.

c. Why is the answer to part *a* so much smaller than the answer to part *b*?

23. Cholesterol Content The average cholesterol content of a certain brand of eggs is 215 milligrams, and the standard deviation is 15 milligrams. Assume the variable is normally distributed.

a. If a single egg is selected, find the probability that the cholesterol content will be greater than 220 milligrams.

b. If a sample of 25 eggs is selected, find the probability that the mean of the sample will be larger than 220 milligrams.

Source: Living Fit.

24. Ages of Proofreaders At a large publishing company, the mean age of proofreaders is 36.2 years, and the standard deviation is 3.7 years. Assume the variable is normally distributed.

a. If a proofreader from the company is randomly selected, find the probability that his or her age will be between 36 and 37.5 years.

b. If a random sample of 15 proofreaders is selected, find the probability that the mean age of the proofreaders in the sample will be between 36 and 37.5 years.

25. TIMSS Test On the Trends in International Mathematics and Science Study (TIMSS) test in a recent year, the United States scored an average of 508 (well below South Korea, 597; Singapore, 593; Hong Kong, 572; and Japan, 570). Suppose that we take a random sample of *n* United States scores and that the population standard deviation is 72. If the probability that the mean of the sample exceeds 520 is 0.0985, what was the sample size?

Source: World Almanac.

Extending the Concepts

For Exercises 26 and 27, check to see whether the correction factor should be used. If so, be sure to include it in the calculations.

26. Life Expectancies In a study of the life expectancy of 500 people in a certain geographic region, the mean age at death was 72.0 years, and the standard deviation was 5.3 years. If a sample of 50 people from this region is selected, find the probability that the mean life expectancy will be less than 70 years.

27. Home Values A study of 800 homeowners in a certain area showed that the average value of the homes was \$82,000, and the standard deviation was \$5000. If 50 homes are for sale, find the probability that the mean of the values of these homes is greater than \$83,500.

28. Breaking Strength of Steel Cable The average breaking strength of a certain brand of steel cable is 2000 pounds, with a standard deviation of 100 pounds.

A sample of 20 cables is selected and tested. Find the sample mean that will cut off the upper 95% of all samples of size 20 taken from the population. Assume the variable is normally distributed.

29. The standard deviation of a variable is 15. If a sample of 100 individuals is selected, compute the standard error of the mean. What size sample is necessary to double the standard error of the mean?

30. In Exercise 29, what size sample is needed to cut the standard error of the mean in half?

6–4 The Normal Approximation to the Binomial Distribution

OBJECTIVE 7

Use the normal approximation to compute probabilities for a binomial variable.

A normal distribution is often used to solve problems that involve the binomial distribution since when n is large (say, 100), the calculations are too difficult to do by hand using the binomial distribution. Recall from Chapter 5 that a binomial distribution has the following characteristics:

1. There must be a fixed number of trials.
2. The outcome of each trial must be independent.
3. Each experiment can have only two outcomes or outcomes that can be reduced to two outcomes.
4. The probability of a success must remain the same for each trial.

Also, recall that a binomial distribution is determined by n (the number of trials) and p (the probability of a success). When p is approximately 0.5, and as n increases, the shape of the binomial distribution becomes similar to that of a normal distribution. The larger n is and the closer p is to 0.5, the more similar the shape of the binomial distribution is to that of a normal distribution.

But when p is close to 0 or 1 and n is relatively small, a normal approximation is inaccurate. As a rule of thumb, statisticians generally agree that a normal approximation should be used only when $n \cdot p$ and $n \cdot q$ are both greater than or equal to 5. (*Note:* $q = 1 - p$.) For example, if p is 0.3 and n is 10, then $np = (10)(0.3) = 3$, and a normal distribution should not be used as an approximation. On the other hand, if $p = 0.5$ and $n = 10$, then $np = (10)(0.5) = 5$ and $nq = (10)(0.5) = 5$, and a normal distribution can be used as an approximation. See Figure 6–38.

In addition to the previous condition of $np \geq 5$ and $nq \geq 5$, a correction for continuity may be used in the normal approximation.

A **correction for continuity** is a correction employed when a continuous distribution is used to approximate a discrete distribution.

The continuity correction means that for any specific value of X, say 8, the boundaries of X in the binomial distribution (in this case, 7.5 to 8.5) must be used. (See Section 1–2.) Hence, when you employ a normal distribution to approximate the binomial, you must use the boundaries of any specific value X as they are shown in the binomial distribution. For example, for $P(X = 8)$, the correction is $P(7.5 < X < 8.5)$. For $P(X \leq 7)$, the correction is $P(X < 7.5)$. For $P(X \geq 3)$, the correction is $P(X > 2.5)$.

Interesting Fact

Of the 12 months, August ranks first in the number of births for Americans.

Students sometimes have difficulty deciding whether to add 0.5 or subtract 0.5 from the data value for the correction factor. Table 6–2 summarizes the different situations.

The formulas for the mean and standard deviation for the binomial distribution are necessary for calculations. They are

$$\mu = n \cdot p \qquad \text{and} \qquad \sigma = \sqrt{n \cdot p \cdot q}$$

FIGURE 6–38
Comparison of the Binomial Distribution and a Normal Distribution

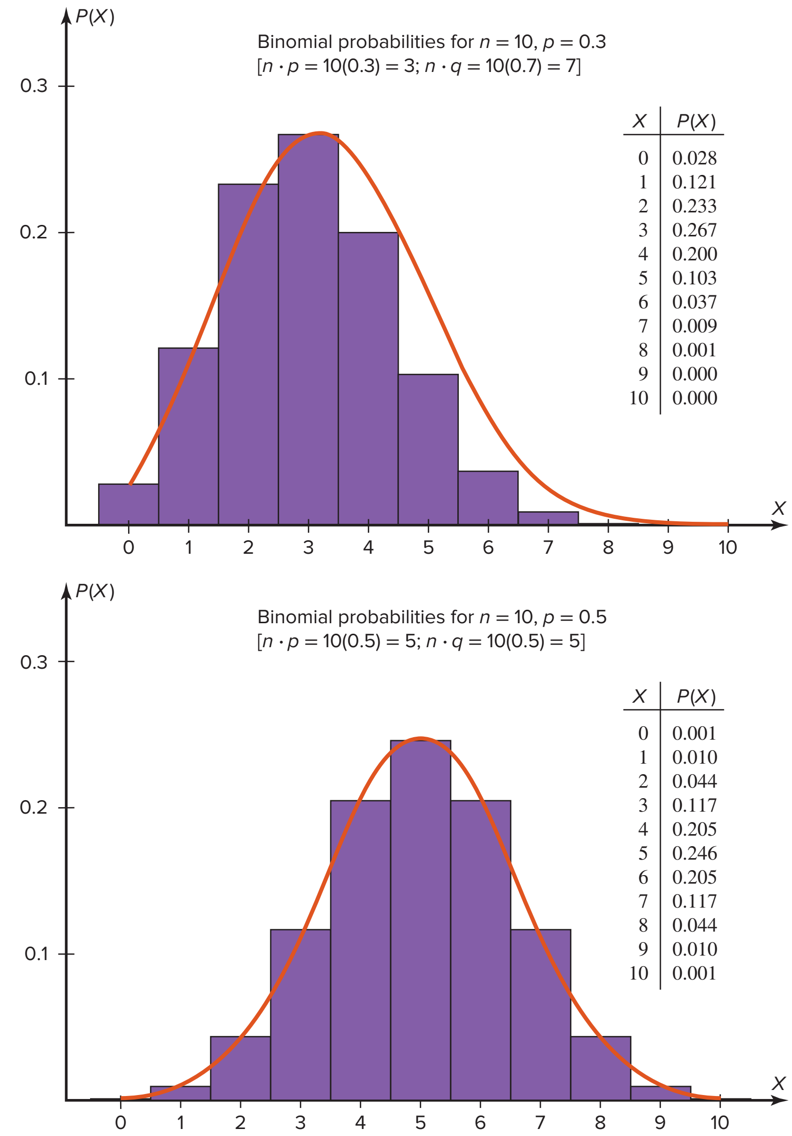


TABLE 6–2 Summary of the Normal Approximation to the Binomial Distribution

Binomial	**Normal**
When finding:	Use:
1. $P(X = a)$	$P(a - 0.5 < X < a + 0.5)$
2. $P(X \geq a)$	$P(X > a - 0.5)$
3. $P(X > a)$	$P(X > a + 0.5)$
4. $P(X \leq a)$	$P(X < a + 0.5)$
5. $P(X < a)$	$P(X < a - 0.5)$

For all cases, $\mu = n \cdot p$, $\sigma = \sqrt{n \cdot p \cdot q}$, $n \cdot p \geq 5$, and $n \cdot q \geq 5$.

The steps for using the normal distribution to approximate the binomial distribution are shown in this Procedure Table.

Procedure Table

Procedure for the Normal Approximation to the Binomial Distribution

Step 1 Check to see whether the normal approximation can be used.

Step 2 Find the mean μ and the standard deviation σ.

Step 3 Write the problem in probability notation, using X.

Step 4 Rewrite the problem by using the continuity correction factor, and show the corresponding area under the normal distribution.

Step 5 Find the corresponding z values.

Step 6 Find the solution.

EXAMPLE 6–16 Falling Asleep While Driving

It has been reported that last year 25% of drivers have fallen asleep at the wheel. If 200 drivers are selected at random, find the probability that 62 will say that they have fallen asleep while driving.

Source: National Sleep Foundation.

SOLUTION

Here $p = 0.25$, $q = 0.75$, and $n = 200$.

Step 1 Check to see whether a normal approximation can be used.

$$np = (200)(0.25) = 50 \qquad nq = (200)(0.75) = 150$$

Since $nq \geq 5$ and $nq \geq 5$, the normal distribution can be used.

Step 2 Find the mean and standard deviation.

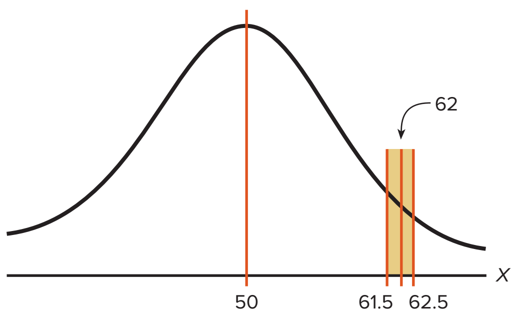
$$\mu = np = (200)(0.25) = 50$$

$$\sigma = \sqrt{npq} = \sqrt{(200)(0.25)(0.75)} = \sqrt{37.5} \approx 6.12$$

Step 3 Write the problem in probability notation: $P(X = 62)$.

Step 4 Rewrite the problem by using the continuity connection factor. See approximation 1 in Table 6–2: $P(62 - 0.5 < X < 62 + 0.5) = P(61.5 < X < 62.5)$. Show the corresponding area under the normal distribution curve. See Figure 6–39.

FIGURE 6–39
Area Under a Normal Curve and X Values for Example 6–16



Step 5 Find the corresponding z values. Since 72 represents any value between 71.5 and 72.5, find both z values.

$$z_1 = \frac{61.5 - 50}{6.12} \approx 1.88$$

$$z_2 = \frac{62.5 - 50}{6.12} \approx 2.04$$

Step 6 The area to the left of $z = 1.88$ is 0.9699, and the area to the left of $z = 2.04$ is 0.9793. The area between the two z values is $0.9793 - 0.9699 = 0.0094$, or about 0.94%. (It is close to 1%.) Hence, the probability that exactly 62 people say that they fell asleep at the wheel while driving is 0.94%.

EXAMPLE 6–17 Ragweed Allergies

Ten percent of Americans are allergic to ragweed. If a random sample of 200 people is selected, find the probability that 10 or more will be allergic to ragweed.

SOLUTION

Step 1 Check to see whether the normal approximate can be used.

Here $p = 0.10$, $q = 0.90$, and $n = 200$. Since $np = (200)(0.10) = 20$ and $nq = (200)(0.90) = 180$, the normal approximation can be used.

Step 2 Find the mean and the standard deviation.

$$\mu = np = (200)(0.10) = 20$$

$$\sigma = \sqrt{npq} = \sqrt{(200)(0.10)(0.90)} = \sqrt{18} \approx 4.24$$

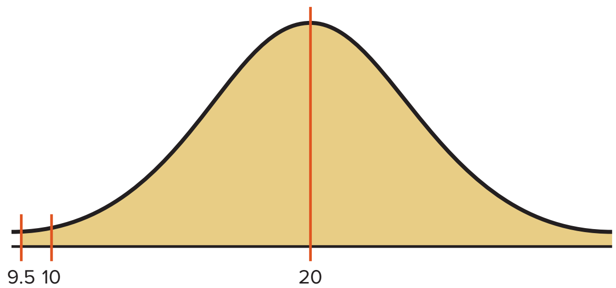
Step 3 Write the problem in probability notation, using X.

$$P(X \geq 10)$$

Step 4 Rewrite the problem, using the continuity correction factor, and show the corresponding area under the normal distribution.

See approximation number 2 in Table 6–2: $P(X > 10 - 0.5) = P(X > 9.5)$. The desired area is shown in Figure 6–40.

FIGURE 6–40
Area Under a Normal Curve and X Value for Example 6–17



Step 5 Find the corresponding z values.

Since the problem is to find the probability of 10 or more positive responses, a normal distribution graph is as shown in Figure 6–40.

The z value is

$$z = \frac{9.5 - 20}{4.24} = -2.48$$

Step 6 Find the solution.

The area to the left of $z = -2.48$ is 0.0066. Hence, the area to the right of $z = -2.48$ is $1.0000 - 0.0066 = 0.9934$, or 99.34%.

It can be concluded, then, that in a random sample of 200 Americans the probability of 10 or more Americans being allergic to ragweed is 99.34%.

EXAMPLE 6–18 Batting Averages

If a baseball player's batting average is 0.320 (32%), find the probability that the player will get at most 26 hits in 100 times at bat.

SOLUTION

Step 1 Check to see whether the normal approximate can be used.

Here, $p = 0.32$, $q = 0.68$, and $n = 100$.

Since $np = (100)(0.320) = 32$ and $nq = (100)(0.680) = 68$, the normal distribution can be used to approximate the binomial distribution.

Step 2 Find the mean and the standard deviation.

$$\mu = np = (100)(0.320) = 32$$

$$\sigma = \sqrt{npq} = \sqrt{(100)(0.32)(0.68)} \approx \sqrt{21.76} \approx 4.66$$

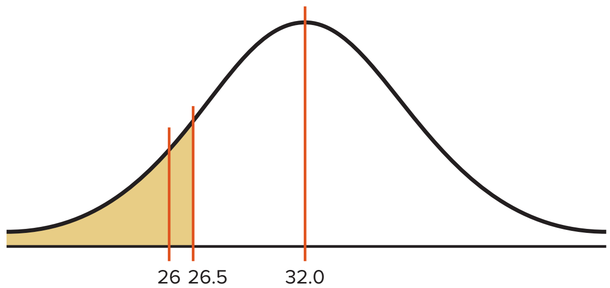
Step 3 Write the problem in probability notation, using X.

$$P(X \leq 26)$$

Step 4 Rewrite the problem using the continuity correction factor and show the corresponding area under the normal distribution.

See approximation number 4 in Table 6–2: $P(X < 26 + 0.5) = P(X < 26.5)$. The desired area is shown in Figure 6–41.

FIGURE 6–41
Area Under a Normal Curve for Example 6–18



Step 5 Find the corresponding z values.

The z value is

$$z = \frac{26.5 - 32}{4.66} \approx -1.18$$

Step 6 Find the solution.

The area to the left of $z = -1.18$ is 0.1190. Hence, the probability is 0.1190, or 11.9%.

It can be concluded that the probability of a player getting at most 26 hits in 100 times at bat is 11.9%.

The closeness of the normal approximation is shown in Example 6–19.

EXAMPLE 6–19 Binomial versus Normal Approximation

When $n = 10$ and $p = 0.5$, use the binomial distribution table (Table B in Appendix A) to find the probability that $X = 6$. Then use the normal approximation to find the probability that $X = 6$.

SOLUTION

From Table B, for $n = 10$, $p = 0.5$, and $X = 6$, the probability is 0.205.
For a normal approximation,

$$\mu = np = (10)(0.5) = 5$$

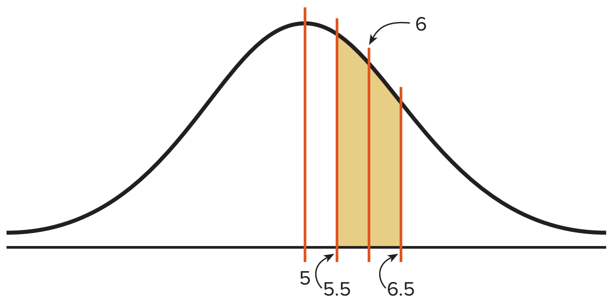
$$\sigma = \sqrt{npq} = \sqrt{(10)(0.5)(0.5)} \approx 1.58$$

Now, $X = 6$ is represented by the boundaries 5.5 and 6.5. So the z values are

$$z_1 = \frac{6.5 - 5}{1.58} \approx 0.95 \qquad z_2 = \frac{5.5 - 5}{1.58} \approx 0.32$$

The corresponding area for 0.95 is 0.8289, and the corresponding area for 0.32 is 0.6255. The area between the two z values of 0.95 and 0.32 is 0.8289 − 0.6255 = 0.2034, which is very close to the binomial table value of 0.205. See Figure 6–42.

FIGURE 6–42
Area Under a Normal Curve for Example 6–19



The normal approximation also can be used to approximate other distributions, such as the Poisson distribution (see Table C in Appendix C).

Applying the Concepts 6–4

Mountain Climbing Safety

Assume one of your favorite activities is mountain climbing. When you go mountain climbing, you have several safety devices to keep you from falling. You notice that attached to one of your safety hooks is a reliability rating of 97%. You estimate that throughout the next year you will be using this device about 100 times. Answer the following questions.

1. Does a reliability rating of 97% mean that there is a 97% chance that the device will not fail any of the 100 times?
2. What is the probability of at least one failure?
3. What is the complement of this event?
4. Can this be considered a binomial experiment?
5. Can you use the binomial probability formula? Why or why not?
6. Find the probability of at least two failures.
7. Can you use a normal distribution to accurately approximate the binomial distribution? Explain why or why not.
8. Is correction for continuity needed?
9. How much safer would it be to use a second safety hook independent of the first?

See page 368 for the answers.

Exercises 6–4

1. Explain why a normal distribution can be used as an approximation to a binomial distribution.

2. What conditions must be met to use the normal distribution to approximate the binomial distribution?

3. Why is a correction for continuity necessary?

4. When is the normal distribution not a good approximation for the binomial distribution?

5. Use the normal approximation to the binomial to find the probabilities for the specific value(s) of X.
   a. $n = 30, p = 0.5, X = 18$
   b. $n = 50, p = 0.8, X = 44$
   c. $n = 100, p = 0.1, X = 12$

6. Use the normal approximation to find the probabilities for the specific value(s) of X.
   a. $n = 10, p = 0.5, X \geq 7$
   b. $n = 20, p = 0.7, X \leq 12$
   c. $n = 50, p = 0.6, X \leq 40$

7. Check each binomial distribution to see whether it can be approximated by a normal distribution (i.e., are $np \geq 5$ and $nq \geq 5$?).
   a. $n = 20, p = 0.5$
   b. $n = 10, p = 0.6$
   c. $n = 40, p = 0.9$

8. Check each binomial distribution to see whether it can be approximated by a normal distribution (i.e., are $np \geq 5$ and $nq \geq 5$?).
   a. $n = 50, p = 0.2$
   b. $n = 30, p = 0.8$
   c. $n = 20, p = 0.85$

9. **Single Americans** In a recent year, about 22% of Americans 18 years and older are single. What is the probability that in a random sample of 200 Americans 18 or older more than 30 are single?
   Source: U.S. Department of Commerce

10. **School Enrollment** Of all 3- to 5-year-old children, 56% are enrolled in school. If a sample of 500 such children is randomly selected, find the probability that at least 250 will be enrolled in school.
    Source: Statistical Abstract of the United States.

11. **Home Ownership** In a recent year, the rate of U.S. home ownership was 65.9%. Choose a random sample of 120 households across the United States. What is the probability that 65 to 85 (inclusive) of them live in homes that they own?
    Source: World Almanac.

12. **Mail Order** A mail order company has an 8% success rate. If it mails advertisements to 600 people, find the probability of getting fewer than 40 sales.

13. **Small Business Owners** Seventy-six percent of small business owners do not have a college degree. If a random sample of 60 small business owners is selected, find the probability that exactly 48 will not have a college degree.
    Source: Business Week.

14. **Selected Technologies** According to the World Almanac, 72% of households own smartphones. If a random sample of 180 households is selected, what is the probability that more than 115 but fewer than 125 have a smartphone?
    Source: World Almanac.

15. **Back Injuries** Twenty-two percent of work injuries are back injuries. If 400 work-injured people are selected at random, find the probability that 92 or fewer have back injuries.
    Source: The World Almanac

16. **Population of College Cities** College students often make up a substantial portion of the population of college cities and towns. State College, Pennsylvania, ranks first with 71.1% of its population made up of college students. What is the probability that in a random sample of 150 people from State College, more than 50 are not college students?
    Source: www.infoplease.com

17. **Mistakes in Restaurant Bills** About 12.5% of restaurant bills are incorrect. If 200 bills are selected at random, find the probability that at least 22 will contain an error. Is this likely or unlikely to occur?
    Source: Harper's Index.

18. **Internet Browsers** The top web browser in 2015 was Chrome with 51.74% of the market. In a random sample of 250 people, what is the probability that fewer than 110 did not use Chrome?
    Source: New York Times Almanac.

19. **Female Americans Who Have Completed 4 Years of College** The percentage of female Americans 25 years old and older who have completed 4 years of college or more is 26.1. In a random sample of 200 American women who are at least 25, what is the probability that at most 50 have completed 4 years of college or more?
    Source: New York Times Almanac.

20. **Residences of U.S. Citizens** According to the U.S. Census, 67.5% of the U.S. population were born in their state of residence. In a random sample of

200 Americans, what is the probability that fewer than 125 were born in their state of residence?

Source: www.census.gov

21. **Elementary School Teachers** Women comprise 80.3% of all elementary school teachers. In a random sample of 300 elementary teachers, what is the probability that less than three-fourths are women?

Source: New York Times Almanac.

22. **Parking Lot Construction** The mayor of a small town estimates that 35% of the residents in the town favor the construction of a municipal parking lot. If there are 350 people at a town meeting, find the probability that at least 100 favor construction of the parking lot. Based on your answer, is it likely that 100 or more people would favor the parking lot?

Extending the Concepts

23. Recall that for use of a normal distribution as an approximation to the binomial distribution, the conditions $np \geq 5$ and $nq \geq 5$ must be met. For each given probability, compute the minimum sample size needed for use of the normal approximation.

a. $p = 0.1$
b. $p = 0.3$
c. $p = 0.5$
d. $p = 0.8$
e. $p = 0.9$

Summary

- A normal distribution can be used to describe a variety of variables, such as heights, weights, and temperatures. A normal distribution is bell-shaped, unimodal, symmetric, and continuous; its mean, median, and mode are equal. Since each normally distributed variable has its own distribution with mean μ and standard deviation σ, mathematicians use the standard normal distribution, which has a mean of 0 and a standard deviation of 1. Other approximately normally distributed variables can be transformed to the standard normal distribution with the formula $z = (X - \mu)/\sigma$. (6–1)
- A normal distribution can be used to solve a variety of problems in which the variables are approximately normally distributed. (6–2)
- A sampling distribution of sample means is a distribution using the means computed from all possible random samples of a specific size taken from a population. The difference between a sample measure and the corresponding population measure is due to what is called *sampling error*. The mean of the sample means will be the same as the population mean. The standard deviation of the sample means will be equal to the population standard deviation divided by the square root of the sample size. The central limit theorem states that as the sample size increases without limit, the shape of the distribution of the sample means taken with replacement from a population will approach that of a normal distribution. (6–3)
- A normal distribution can be used to approximate other distributions, such as a binomial distribution. For a normal distribution to be used as an approximation, the conditions $np \geq 5$ and $nq \geq 5$ must be met. Also, a correction for continuity may be used for more accurate results. (6–4)

Important Terms

central limit theorem 346

correction for continuity 354

negatively or left-skewed distribution 315

normal distribution 313

positively or right-skewed distribution 315

sampling distribution of sample means 344

sampling error 344

standard error of the mean 346

standard normal distribution 315

symmetric distribution 314

z value (z score) 316

Important Formulas

Formula for the z score (or standard score):

$$z = \frac{X - \mu}{\sigma}$$

Formula for finding a specific data value:

$$X = z \cdot \sigma + \mu$$

Formula for the mean of the sample means:

$$\mu_{\overline{X}} = \mu$$

Formula for the standard error of the mean:

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

Formula for the z value for the central limit theorem:

$$z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

Formulas for the mean and standard deviation for the binomial distribution:

$$\mu = n \cdot p \qquad \sigma = \sqrt{n \cdot p \cdot q}$$

Review Exercises

Section 6–1

1. Find the area under the standard normal distribution curve for each.
   a. Between $z = 0$ and $z = 2.06$
   b. Between $z = 0$ and $z = 0.53$
   c. Between $z = 1.26$ and 1.74
   d. Between $z = -1.02$ and $z = 1.63$
   e. Between $z = -0.07$ and $z = 0.49$

2. Find the area under the standard normal distribution for each.
   a. Between $z = 1.10$ and $z = -1.80$
   b. To the right of $z = 1.99$
   c. To the right of $z = -1.36$
   d. To the left of $z = -2.09$
   e. To the left of $z = 1.68$

3. Using the standard normal distribution, find each probability.
   a. $P(0 < z < 2.23)$
   b. $P(-1.75 < z < 0)$
   c. $P(-1.48 < z < 1.68)$
   d. $P(1.22 < z < 1.77)$
   e. $P(-2.31 < z < 0.32)$

4. Using the standard normal distribution, find each probability.
   a. $P(z > 1.66)$
   b. $P(z < -2.03)$
   c. $P(z > -1.19)$
   d. $P(z < 1.93)$
   e. $P(z > -1.77)$

Section 6–2

5. **Per Capita Spending on Health Care** The average per capita spending on health care in the United States is \$5274. If the standard deviation is \$600 and the distribution of health care spending is approximately normal, what is the probability that a randomly selected person spends more than \$6000? Find the limits of the middle 50% of individual health care expenditures.
   Source: World Almanac.

6. **Salaries for Actuaries** The average salary for graduates entering the actuarial field is \$63,000. If the salaries are normally distributed with a standard deviation of \$5000, find the probability that
   a. An individual graduate will have a salary over \$68,000.
   b. A group of nine graduates will have a group average over \$68,000.
   Source: www.payscale.com

7. **Commuter Train Passengers** On a certain run of a commuter train, the average number of passengers is 476 and the standard deviation is 22. Assume the variable is normally distributed. If the train makes the run, find the probability that the number of passengers will be
   a. Between 476 and 500 passengers
   b. Fewer than 450 passengers
   c. More than 510 passengers

8. **Monthly Spending for Paging and Messaging Services** The average individual monthly spending in the United States for paging and messaging services is \$10.15. If the standard deviation is \$2.45 and the amounts are normally distributed, what is the probability that a randomly selected user of these services pays more than \$15.00 per month? Between \$12.00 and \$14.00 per month?
   Source: New York Times Almanac.

9. **Cost of Smartphone Repair** The average cost of repairing an smartphone is \$120 with a standard deviation of \$10.50. The costs are normally distributed. If 15% of the costs are considered excessive, find the cost in dollars that would be considered excessive.

10. **Slot Machine Earnings** The average amount a slot machine makes per month is \$8000. This is after payouts. Assume the earnings are normally distributed with a standard deviation of \$750. Find the amount of earnings if 15% is considered too low.

11. **Private Four-Year College Enrollment** A random sample of enrollments in Pennsylvania's private four-year colleges is listed here. Check for normality.

1350	1886	1743	1290	1767
2067	1118	3980	1773	4605
1445	3883	1486	980	1217
3587				

Source: New York Times Almanac.

12. **Heights of Active Volcanoes** The heights (in feet above sea level) of a random sample of the world's active volcanoes are shown here. Check for normality.

13,435	5,135	11,339	12,224	7,470
9,482	12,381	7,674	5,223	5,631
3,566	7,113	5,850	5,679	15,584
5,587	8,077	9,550	8,064	2,686
5,250	6,351	4,594	2,621	9,348
6,013	2,398	5,658	2,145	3,038

Source: New York Times Almanac.

Section 6–3

13. **Confectionary Products** Americans ate an average of 25.7 pounds of confectionary products each last year and spent an average of \$61.50 per person doing so. If the standard deviation for consumption is 3.75 pounds and the standard deviation for the amount spent is \$5.89, find the following:
    a. The probability that the sample mean confectionary consumption for a random sample of 40 American consumers was greater than 27 pounds
    b. The probability that for a random sample of 50, the sample mean for confectionary spending exceeded \$60.00

Source: www.census.gov

14. **Average Precipitation** For the first 7 months of the year, the average precipitation in Toledo, Ohio, is 19.32 inches. If the average precipitation is normally distributed with a standard deviation of 2.44 inches, find these probabilities.
    a. A randomly selected year will have precipitation greater than 18 inches for the first 7 months.
    b. Five randomly selected years will have an average precipitation greater than 18 inches for the first 7 months.

Source: Toledo Blade.

15. **Sodium in Frozen Food** The average number of milligrams (mg) of sodium in a certain brand of low-salt microwave frozen dinners is 660 mg, and the standard deviation is 35 mg. Assume the variable is normally distributed.
    a. If a single dinner is selected, find the probability that the sodium content will be more than 670 mg.
    b. If a sample of 10 dinners is selected, find the probability that the mean of the sample will be larger than 670 mg.
    c. Why is the probability for part *a* greater than that for part *b*?

16. **Wireless Sound System Lifetimes** A recent study of the life span of wireless sound systems found the average to be 3.7 years with a standard deviation of 0.6 year. If a random sample of 32 people who own wireless sound systems is selected, find the probability that the mean lifetime of the sample will be less than 3.4 years. If the sample mean is less than 3.4 years, would you consider that 3.7 years might be incorrect?

Section 6–4

17. **Retirement Income** Of the total population of American households, including older Americans and perhaps some not so old, 17.3% receive retirement income. In a random sample of 120 households, what is the probability that more than 20 households but fewer than 35 households receive a retirement income?

Source: www.bls.gov

18. **Slot Machines** The probability of winning on a slot machine is 5%. If a person plays the machine 500 times, find the probability of winning 30 times. Use the normal approximation to the binomial distribution.

19. **Multiple-Job Holders** According to the government, 5.3% of those employed are multiple-job holders. In a random sample of 150 people who are employed, what is the probability that fewer than 10 hold multiple jobs? What is the probability that more than 50 are not multiple-job holders?

Source: www.bls.gov

20. **Enrollment in Personal Finance Course** In a large university, 30% of the incoming first-year students elect to enroll in a personal finance course offered by the university. Find the probability that of 800 randomly selected incoming first-year students, at least 260 have elected to enroll in the course.

21. **U.S. Population** Of the total U.S. population, 37% live in the South. If 200 U.S. residents are selected at random, find the probability that at least 80 live in the South.

Source: Statistical Abstract of the United States.

22. **Larceny-Thefts** Excluding motor vehicle thefts, 26% of all larceny-thefts involved items taken from motor vehicles. Local police forces are trying to help the situation with their "Put your junk in the trunk!" campaign. Consider a random sample of 60 larceny-thefts. What is the probability that 20 or more were items stolen from motor vehicles?

Source: World Almanac.

STATISTICS TODAY

What Is Normal?—Revisited

Many of the variables measured in medical tests—blood pressure, triglyceride level, etc.—are approximately normally distributed for the majority of the population in the United States. Thus, researchers can find the mean and standard deviation of these variables. Then, using these two measures along with the z values, they can find normal intervals for healthy individuals. For example, 95% of the systolic blood pressures of healthy individuals fall within 2 standard deviations of the mean. If an individual's pressure is outside the determined normal range (either above or below), the physician will look for a possible cause and prescribe treatment if necessary.

Chapter Quiz

Determine whether each statement is true or false. If the statement is false, explain why.

1. The total area under a normal distribution is infinite.
2. The standard normal distribution is a continuous distribution.
3. All variables that are approximately normally distributed can be transformed to standard normal variables.
4. The z value corresponding to a number below the mean is always negative.
5. The area under the standard normal distribution to the left of $z = 0$ is negative.
6. The central limit theorem applies to means of samples selected from different populations.

Select the best answer.

7. The mean of the standard normal distribution is
   a. 0
   b. 1
   c. 100
   d. Variable
8. Approximately what percentage of normally distributed data values will fall within 1 standard deviation above or below the mean?
   a. 68%
   b. 95%
   c. 99.7%
   d. Variable
9. Which is not a property of the standard normal distribution?
   a. It's symmetric about the mean.
   b. It's uniform.
   c. It's bell-shaped.
   d. It's unimodal.
10. When a distribution is positively skewed, the relationship of the mean, median, and mode from left to right will be
    a. Mean, median, mode
    b. Mode, median, mean
    c. Median, mode, mean
    d. Mean, mode, median
11. The standard deviation of all possible sample means equals
    a. The population standard deviation
    b. The population standard deviation divided by the population mean
    c. The population standard deviation divided by the square root of the sample size
    d. The square root of the population standard deviation

Complete the following statements with the best answer.

12. When one is using the standard normal distribution, $P(z < 0) =$ ________.
13. The difference between a sample mean and a population mean is due to ________.
14. The mean of the sample means equals ________.
15. The standard deviation of all possible sample means is called the ________.
16. The normal distribution can be used to approximate the binomial distribution when $n \cdot p$ and $n \cdot q$ are both greater than or equal to ________.
17. The correction factor for the central limit theorem should be used when the sample size is greater than ________ of the size of the population.
18. Find the area under the standard normal distribution for each.
    a. Between 0 and 1.50
    b. Between 0 and −1.25
    c. Between 1.56 and 1.96
    d. Between −1.20 and −2.25
    e. Between −0.06 and 0.73
    f. Between 1.10 and −1.80
    g. To the right of $z = 1.75$
    h. To the right of $z = -1.28$
    i. To the left of $z = -2.12$
    j. To the left of $z = 1.36$

19. Using the standard normal distribution, find each probability.
a. $P(0 < z < 2.16)$
b. $P(-1.87 < z < 0)$
c. $P(-1.63 < z < 2.17)$
d. $P(1.72 < z < 1.98)$
e. $P(-2.17 < z < 0.71)$
f. $P(z > 1.77)$
g. $P(z < -2.37)$
h. $P(z > -1.73)$
i. $P(z < 2.03)$
j. $P(z > -1.02)$

20. Amount of Rain in a City The average amount of rain per year in Greenville is 49 inches. The standard deviation is 8 inches. Find the probability that next year Greenville will receive the following amount of rainfall. Assume the variable is normally distributed.
a. At most 55 inches of rain
b. At least 62 inches of rain
c. Between 46 and 54 inches of rain
d. How many inches of rain would you consider to be an extremely wet year?

21. Heights of People The average height of a certain age group of people is 53 inches. The standard deviation is 4 inches. If the variable is normally distributed, find the probability that a selected individual's height will be
a. Greater than 59 inches
b. Less than 45 inches
c. Between 50 and 55 inches
d. Between 58 and 62 inches

22. Sports Drink Consumption The average number of gallons of sports drinks consumed by the football team during a game is 20, with a standard deviation of 3 gallons. Assume the variable is normally distributed. When a game is played, find the probability of using
a. Between 20 and 25 gallons
b. Less than 19 gallons
c. More than 21 gallons
d. Between 26 and 28 gallons

23. Years to Complete a Graduate Program The average number of years a person takes to complete a graduate degree program is 3. The standard deviation is 4 months. Assume the variable is normally distributed. If an individual enrolls in the program, find the probability that it will take
a. More than 4 years to complete the program
b. Less than 3 years to complete the program
c. Between 3.8 and 4.5 years to complete the program
d. Between 2.5 and 3.1 years to complete the program

24. Passengers on a Bus On the daily run of an express bus, the average number of passengers is 48. The standard deviation is 3. Assume the variable is normally distributed. Find the probability that the bus will have
a. Between 36 and 40 passengers
b. Fewer than 42 passengers
c. More than 48 passengers
d. Between 43 and 47 passengers

25. Thickness of Library Books The average thickness of books on a library shelf is 8.3 centimeters. The standard deviation is 0.6 centimeter. If 20% of the books are oversized, find the minimum thickness of the oversized books on the library shelf. Assume the variable is normally distributed.

26. Membership in an Organization Membership in an elite organization requires a test score in the upper 30% range. If $\mu = 115$ and $\sigma = 12$, find the lowest acceptable score that would enable a candidate to apply for membership. Assume the variable is normally distributed.

27. Repair Cost for Microwave Ovens The average repair cost of a microwave oven is \$55, with a standard deviation of \$8. The costs are normally distributed. If 12 ovens are repaired, find the probability that the mean of the repair bills will be greater than \$60.

28. Electric Bills The average electric bill in a residential area is \$72 for the month of April. The standard deviation is \$6. If the amounts of the electric bills are normally distributed, find the probability that the mean of the bill for 15 residents will be less than \$75.

29. Sleep Survey According to a recent survey, 38% of Americans get 6 hours or less of sleep each night. If 25 people are selected, find the probability that 14 or more people will get 6 hours or less of sleep each night. Does this number seem likely?

Source: Amazing Almanac.

30. Unemployment If 8% of all people in a certain geographic region are unemployed, find the probability that in a sample of 200 people, fewer than 10 people are unemployed.

31. Household Online Connection The percentage of U.S. households that have online connections is 78%. In a random sample of 420 households, what is the probability that fewer than 315 have online connections?

Source: World Almanac.

32. Computer Ownership Fifty-three percent of U.S. households have a personal computer. In a random sample of 250 households, what is the probability that fewer than 120 have a PC?

Source: New York Times Almanac.

33. Calories in Fast-Food Sandwiches The number of calories contained in a selection of fast-food sandwiches is shown here. Check for normality.

390	405	580	300	320
540	225	720	470	560
535	660	530	290	440
390	675	530	1010	450
320	460	290	340	610
430	530			

Source: The Doctor's Pocket Calorie, Fat, and Carbohydrate Counter.

34. GMAT Scores The average GMAT scores for the top-30 ranked graduate schools of business are listed here. Check for normality.

718	703	703	703	700	690	695	705	690	688
676	681	689	686	691	669	674	652	680	670
651	651	637	662	641	645	645	642	660	636

Source: U.S. News & World Report Best Graduate Schools.

Critical Thinking Challenges

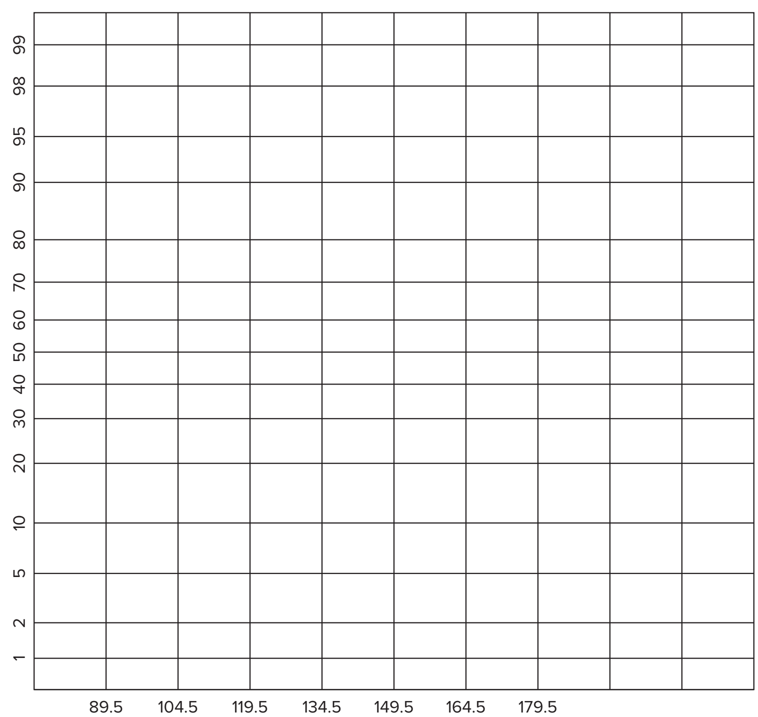
Sometimes a researcher must decide whether a variable is normally distributed. There are several ways to do this. One simple but very subjective method uses special graph paper, which is called *normal probability paper*. For the distribution of systolic blood pressure readings given in Chapter 3 of the text, the following method can be used:

1. Make a table, as shown.

Boundaries	Frequency	Cumulative frequency	Cumulative percent frequency
89.5–104.5	24		
104.5–119.5	62		
119.5–134.5	72		
134.5–149.5	26		
149.5–164.5	12		
164.5–179.5	4		
	200		

2. Find the cumulative frequencies for each class, and place the results in the third column.
3. Find the cumulative percents for each class by dividing each cumulative frequency by 200 (the total frequencies) and multiplying by 100%. (For the first class, it would be $24/200 \times 100\% = 12\%$.) Place these values in the last column.
4. Using the normal probability paper shown in Table 6–3, label the x axis with the class boundaries as shown and plot the percents.
5. If the points fall approximately in a straight line, it can be concluded that the distribution is normal. Do you feel that this distribution is approximately normal? Explain your answer.
6. To find an approximation of the mean or median, draw a horizontal line from the 50% point on the y axis over

TABLE 6–3 Normal Probability Paper



to the curve and then a vertical line down to the x axis. Compare this approximation of the mean with the computed mean.

7. To find an approximation of the standard deviation, locate the values on the x axis that correspond to the 16 and 84% values on the y axis. Subtract these two values and divide the result by 2. Compare this approximate standard deviation to the computed standard deviation.

8. Explain why the method used in step 7 works.

Data Projects

1. **Business and Finance** Use the data collected in data project 1 of Chapter 2 regarding earnings per share to complete this problem. Use the mean and standard deviation computed in data project 1 of Chapter 3 as estimates for the population parameters. What value separates the top 5% of stocks from the others?

2. **Sports and Leisure** Find the mean and standard deviation for the batting average for a player in the most recently completed MLB season. What batting average would separate the top 5% of all hitters from the rest? What is the probability that a randomly selected player bats over 0.300? What is the probability that a team of 25 players has a mean that is above 0.275?

3. **Technology** Use the data collected in data project 3 of Chapter 2 regarding song lengths. If the sample estimates for mean and standard deviation are used as replacements for the population parameters for this data set, what song length separates the bottom 5% and top 5% from the other values?

4. **Health and Wellness** Use the data regarding heart rates collected in data project 6 of Chapter 2 for this problem. Use the sample mean and standard deviation as estimates of the population parameters. For the before-exercise data, what heart rate separates the top 10% from the other values? For the after-exercise data, what heart rate separates the bottom 10% from the other values? If a student were selected at random, what would be the probability of her or his mean heart rate before exercise being less than 72? If 25 students were selected at random, what would be the probability that their mean heart rate before exercise was less than 72?

5. **Politics and Economics** Collect data regarding Math SAT scores to complete this problem. What are the mean and standard deviation for statewide Math SAT scores? What SAT score separates the bottom 10% of states from the others? What is the probability that a randomly selected state has a statewide SAT score above 500?

6. **Formulas** Confirm the two formulas hold true for the central limit theorem for the population containing the elements {1, 5, 10}. First, compute the population mean and standard deviation for the data set. Next, create a list of all 9 of the possible two-element samples that can be created with replacement: {1, 1}, {1, 5}, etc. For each of the 9 compute the sample mean. Now find the mean of the sample means. Does it equal the population mean? Compute the standard deviation of the sample means. Does it equal the population standard deviation, divided by the square root of n?

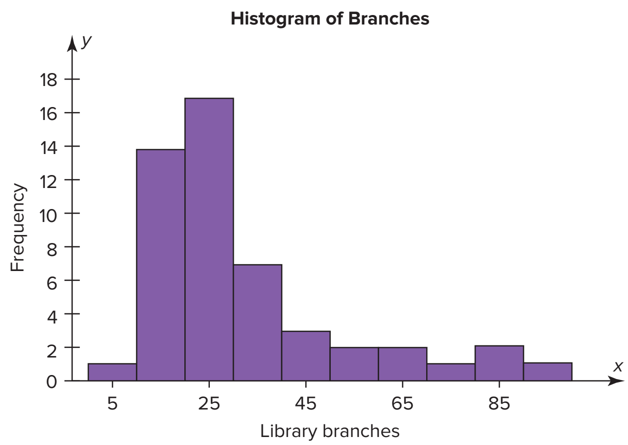
Answers to Applying the Concepts

Section 6–1 Assessing Normality

1. Answers will vary. One possible frequency distribution is the following:

Limits	Frequency
0–9	1
10–19	14
20–29	17
30–39	7
40–49	3
50–59	2
60–69	2
70–79	1
80–89	2
90–99	1

2. Answers will vary according to the frequency distribution in question 1. This histogram matches the frequency distribution in question 1.



3. The histogram is unimodal and skewed to the right (positively skewed).

4. The distribution does not appear to be normal.

5. The mean number of branches is $\overline{X} = 31.4$, and the standard deviation is $s = 20.6$.
6. Of the data values, 80% fall within 1 standard deviation of the mean (between 10.8 and 52).
7. Of the data values, 92% fall within 2 standard deviations of the mean (between 0 and 72.6).
8. Of the data values, 98% fall within 3 standard deviations of the mean (between 0 and 93.2).
9. My values in questions 6–8 differ from the 68, 95, and 100% that we would see in a normal distribution.
10. These values support the conclusion that the distribution of the variable is not normal.

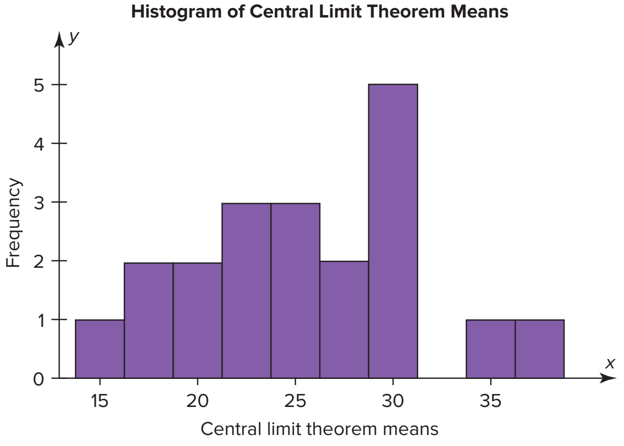
Section 6–2 Smart People

1. $z = \frac{130 - 100}{15} = 2$. The area to the right of 2 in the standard normal table is about 0.0228, so I would expect about 10,000(0.0228) = 228 people in my hometown to qualify for Mensa.
2. It does seem reasonable to continue my quest to start a Mensa chapter in my hometown.
3. Answers will vary. One possible answer would be to randomly call telephone numbers (both home and cell phones) in my hometown, ask to speak to an adult, and ask whether the person would be interested in joining Mensa.
4. To have an Ultra-Mensa club, I would need to find the people in my hometown who have IQs that are at least 2.326 standard deviations above average. This means that I would need to recruit those with IQs that are at least 135:

$$2.326 = \frac{x - 100}{15} \Rightarrow x = 100 + 2.326(15) = 134.89$$

Section 6–3 Times To Travel to School

1. It is very unlikely that we would ever get the same results for any of our random samples. While it is a remote possibility, it is highly unlikely.
2. A good estimate for the population mean would be to find the average of the students' sample means. Similarly, a good estimate for the population standard deviation would be to find the average of the students' sample standard deviations.
3. The distribution appears to be somewhat left-skewed (negatively skewed).



4. The mean of the students' means is 25.4, and the standard deviation is 5.8.
5. The distribution of the means is not a sampling distribution, since it represents just 20 of all possible samples of size 30 from the population.
6. The sampling error for student 3 is $18 - 25.4 = -7.4$; the sampling error for student 7 is $26 - 25.4 = +0.6$; the sampling error for student 14 is $29 - 25.4 = +3.6$.
7. The standard deviation for the sample of the 20 means is greater than the standard deviations for each of the individual students. So it is not equal to the standard deviation divided by the square root of the sample size.

Section 6–4 Mountain Climbing Safety

1. A reliability rating of 97% means that, on average, the device will not fail 97% of the time. We do not know how many times it will fail for any particular set of 100 climbs.
2. The probability of at least 1 failure in 100 climbs is $1 - (0.97)^{100} = 1 - 0.0476 = 0.9524$ (about 95%).
3. The complement of the event in question 2 is the event of "no failures in 100 climbs."
4. This can be considered a binomial experiment. We have two outcomes: success and failure. The probability of the equipment working (success) remains constant at 97%. We have 100 independent climbs. And we are counting the number of times the equipment works in these 100 climbs.
5. We could use the binomial probability formula, but it would be very messy computationally.
6. The probability of at least two failures *cannot* be estimated with the normal distribution (see below). So the probability is $1 - [(0.97)^{100} + 100(0.97)^{99}(0.03)] = 1 - 0.1946 = 0.8054$ (about 80.5%).
7. We *should not* use the normal approximation to the binomial since $nq < 5$.
8. If we had used the normal approximation, we would have needed a correction for continuity, since we would have been approximating a discrete distribution with a continuous distribution.
9. Since a second safety hook will be successful or will fail independently of the first safety hook, the probability of failure drops from 3% to (0.03)(0.03) = 0.0009, or 0.09%.